

UNIVERSIDAD AUTÓNOMA DE MADRID
FACULTAD DE CIENCIAS
DEPARTAMENTO DE BIOLOGÍA MOLECULAR Y CELULAR

**Caracterización estructural y
funcional del complejo THO
de *Saccharomyces cerevisiae***

TESIS DOCTORAL

ÁLVARO PEÑA CONDE
Octubre, 2010

UNIVERSIDAD AUTÓNOMA DE MADRID
FACULTAD DE CIENCIAS
DEPARTAMENTO DE BIOLOGÍA MOLECULAR Y CELULAR

Caracterización estructural y funcional del complejo THO de *Saccharomyces cerevisiae*

El trabajo presentado en esta Memoria ha sido realizado por Álvaro Peña Conde en la unidad de Microscopia Electrónica del Departamento de Estructura de Macromoléculas del Centro Nacional de Biotecnología (CSIC), bajo la dirección de los doctores José María Valpuesta y José L. Carrascosa.

La realización de esta Tesis Doctoral ha sido posible gracias a la concesión de una beca predoctoral de Formación de Personal Investigador (FPI) del Ministerio de Educación y Ciencia.

A mis padres, Uhte, Vasqui y Ami.

Los adultos siempre necesitan explicaciones
(El Principito).



Agradecimientos

¡Hay tanto que agradecer!. Son muchas las cosas y las personas que deberían figurar en este apartado y quizá tendría que ser tan largo como el resto de la tesis. Seré breve.

Recién salido de la carreta, llegué un día al CNB, ingenuo, disperso y torpe. Allí conocí a la primera persona a la que quiero agradecer este trabajo, mi director de tesis. Por ofrecerme la oportunidad, por confiar en mí, por enseñarme lo que es “hacer ciencia”, por dedicarme tantas veces su tiempo y sus ideas y por apoyarme en todo momento, incluso cuando el viento soplaba en contra. José María Valpuesta, un gran científico al que debo mucho de lo que he aprendido estos años.

Siempre atento, desde su despacho, desde la bandeja del correo electrónico o en un cruce casual en el pasillo, quiero agradecer el apoyo de mi codirector de tesis, José L. Carrascosa. La persona que me ha enseñado que pase lo que pase y sea como sea, una sonrisa siempre lo hace todo mejor. La ciencia empieza en paciencia.

Ambos me han enseñado que saber y saberlo explicar es valer dos veces.

Durante estos años he avanzando pasito a pasito, en lo profesional y en lo personal. Jorge Cuéllar es la persona que ha caminado conmigo, día a día y ha ido escuchando cada uno de mis planteamientos, de mis ideas, de mis locuras. De él me llevo muchas cosas, pero aquí destaco su frase “hay que ser práctico”, todo un antídoto frente a mentes dispersas e impacientes. Los pies en el suelo. Jorge me ha enseñado que responder “sí” o “no” es breve de decir, pero requiere mucho que pensar.

Noe, Gloria, Susana y Mari Carmen fueron las personas que me acogieron y me enseñaron la importancia de tener buenos compañeros, puntos de apoyo. Noe continúa ahora al pie del cañón, con su tesis y siendo todo un ejemplo para los demás. Además de un excelente científico, seguro será también una gran madre. Los compañeros del sótano, el ambiente y la música de cada día. Elena Pascual y Ana Cuervo han contribuido especialmente a que esta tesis sea lo que es. Además se han ofrecido a leerla y comentar sus opiniones, que han sido de gran utilidad en las fases finales del trabajo. Agradezco a

Jaime, Rocío, Dani, Javi Chichón, Marina, José, Srdja, Josué, Sara, Borja, Elías, Silvia, JR, Encarna, Alina, Mar, Ana B., Adriana, María, Ricardo, Roberto, M^a José, M^a Ángeles, Hugo... a los compañeros del SME, Javi Bueno y Cristina Patiño y a los compañeros de la unidad de biocomputación, Roberto Melero, Sjors, Roberto Marabini, Marta, COS, Carmen San Martín... todos aportando su granito de arena de una u otra forma.

Tamibién quiero expresar aquí mi más sincero agradecimiento a la gente que ha contribuido directamente a la realización de este trabajo. Kamil Gewartowski y Andrzej Dziembowski, que me acogieron en su laboratorio durante un gélido invierno de 2009 y me enseñaron otra forma de hacer ciencia (y de beber vodka). A Sunita y a Guillermo Montoya, por su interés, su disponibilidad, sus propuestas y sus ganas de ayudar. A Andrés Aguilera, por compartir sus ideas y sus anticuerpos anti-THO. A Juan Carlos Sánchez y a Sito, por esas interminables horas de predicción bioinformática de modelos estructurales.

Fuera del ámbito científico, estoy muy agradecido a mis padres, que me han enseñado a buscar la solución, a sacar fuerzas de flaqueza, a optimizar las acciones para que el mundo sea mejor. A mi hermano y la Vasqui, un gran apoyo del que va por dentro, mas allá de lo imaginable. Su apoyo es también parte de esta tesis. Agradezco a mi familia por enseñarme a no rendirme cuando las cosas se ponen difíciles, a levantar la cabeza, coger aire y seguir adelante. Merche, Jürgen y Víctor, el incesante ánimo que siempre llega desde el norte. Nada sería igual sin ellos. Y por supuesto, la parte artística de cada día, Rober, Javito Kops, Santi (Killer), Brian, Jon, Josito, Manu, Sergio, Jean Pierre. Los que comparten la pasión por crear, por crecer, por seguir caminando. La ilusión de ser quienes son, es contagiosa.

Aunque sé que piensa que el final es para lo menos importante, he querido dejar este último párrafo para hacer una dedicación muy especial a Ami, la persona que hace que todo tenga sentido. Este trabajo debe mucho a su apoyo, a su confianza, a su refuerzo positivo, a sus ganas de vivir. La sonrisa que lo llena todo de color, la luz que inspira y fortalece. La más luchadora, la valiente, la rebelde, la que siempre encuentra la mejor solución. Mi gran apoyo. Mi copiloto.

Índice

ÍNDICE GENERAL

ABREVIATURAS Y ACRÓNIMOS	1
SUMMARY	7
1. INTRODUCCIÓN	11
1.1. Flujo de información y expresión génica en células eucariotas	13
1.2. Transcripción	14
1.3. Maduración del ARN mensajero	16
1.4. Biogénesis de RNPM	18
1.5. El complejo THO: perspectiva	21
1.6. Componentes del complejo THO de <i>Saccharomyces cerevisiae</i>	25
1.7. Modelo de acción del complejo THO	28
1.8. Microscopía electrónica y procesamiento digital de imágenes	30
1.9. Importancia del estudio estructural del complejo THO	33
2. OBJETIVOS	37
3. MATERIALES Y MÉTODOS	41
3.1. Obtención de muestras	
3.1.1. Inserción de la cola TAP	43
3.1.2. Cultivo de la cepa Tho2-TAP y purificación del complejo	44
3.1.3. Preparación de cepas mutantes	47
3.1.4. Ensayos de electroforesis y <i>western-blot</i>	48
3.1.5. Diseño del anticuerpo y preparación inmunocomplejos	50
3.2. Cálculo estequiométrico	52
3.3. Análisis de la interacción con ácidos nucleicos	52
3.4. Estudio fenotípico	53
3.5. Predicción de estructura terciaria	54
3.6. Microscopía electrónica	
3.6.1. Preparación de muestras de tinción negativa	55
3.6.2. Adquisición de imágenes	57

3.7. Procesamiento de partículas individuales	
3.7.1. Digitalización y cálculo de la Transformada de Fourier	57
3.7.2. Selección de partículas, corrección de CTF y análisis en 2D	57
3.7.3. Reconstrucción tridimensional	
3.7.3.1. Obtención de volúmenes de referencia	60
3.7.3.2. Refinamiento angular iterativo	62
3.7.3.3. Cálculo de resolución	66
4. RESULTADOS	69
4.1. Purificación del complejo a partir de la cepa Tho2-TAP	71
4.2. Identificación de la proteína Tex1 asociada al complejo THO	74
4.3. Reconstrucción tridimensional del complejo THO:Tex1	75
4.4. Purificación del complejo THO a partir de la cepa THO Δ Tex1	82
4.5. Cálculo estequiométrico de los componentes del complejo THO	84
4.6. Reconstrucción tridimensional del complejo THO	85
4.7. Predicción de estructura de la proteína Tex1	88
4.8. Localización de la proteína Tex1 en el complejo THO:Tex1	91
4.9. Determinación de la posición de Hpr1 en el complejo THO	93
4.10. Caracterización de la interacción del complejo THO con el ADN y ARN	95
4.11. Estudio fenotípico de las cepas mutantes en el complejo THO	100
5. DISCUSIÓN	105
5.1. El concepto TREX	107
5.2. El complejo THO y las proteínas WD	111
5.3. Tho2 en la interacción con los ácidos nucleicos	115
5.4. Hpr1 en la biogénesis de RNPm	119
5.5. El complejo THO en la expresión génica	122
6. CONCLUSIONES	133
7. BIBLIOGRAFÍA	137

ÍNDICE DE FIGURAS

INTRODUCCIÓN

Figura 1: Etapas iniciales de la expresión génica en células eucariotas	19
Figura 2: Principales aportaciones en relación al complejo THO en los últimos veinte años	24
Figura 3: Componentes del complejo THO	26
Figura 4: Modelo de acción del complejo THO en <i>Saccharomyces cerevisiae</i>	29
Figura 5: Etapas principales del procesamiento de partículas individuales por ME	32

MATERIALES Y MÉTODOS

Figura 6: Esquema del fragmento TAP y de su inserción en el complejo THO	43
Figura 7: Secuencia de la proteína Tho2 en las cepas Tho2-TAP, Tho2-C3 y sh-Tho2	48
Figura 8: Diseño del anticuerpo anti-Hpr1	51
Figura 9: Preparación de rejillas para microscopía electrónica	56
Figura 10: Proyecciones teóricas y proyecciones experimentales	63

RESULTADOS

Figura 11: Resultados de la purificación a partir de la cepa Tho2-TAP	73
Figura 12: Identificación de Tex1 de <i>S. cerevisiae</i>	74
Figura 13: Observación y selección de partículas del complejo THO:Tex1	76
Figura 14: Análisis bidimensional del complejo THO:Tex1 de <i>S. cerevisiae</i>	77
Figura 15: Evolución del refinamiento en la reconstrucción del complejo THO:Tex1	78
Figura 16: Reconstrucción tridimensional del complejo THO:Tex1	80
Figura 17: Distribución angular de partículas en la reconstrucción del complejo THO:Tex1	82
Figura 18: Purificación del complejo THO a partir de la cepa THO Δ Tex1	83
Figura 19: Análisis del patrón de densidad de las subunidades del complejo THO	84
Figura 20: Procesamiento bidimensional del complejo THO	86
Figura 21: Reconstrucción tridimensional del complejo THO	87
Figura 22: Predicción de la estructura secundaria de la proteína Tex1 de <i>S. cerevisiae</i>	89
Figura 23: Modelo atómico de la región amino-terminal de la proteína Tex1 de <i>S. cerevisiae</i>	90
Figura 24: Localización y ajuste del modelo de Tex1 en la estructura del complejo THO:Tex1	92
Figura 25: Identificación de la proteína Hpr1 en el complejo THO:Tex1	94

Figura 26: Tho2 interacciona con los ácidos nucleicos en presencia y ausencia de Tex1	96
Figura 27: La región carboxi-terminal de Tho2 es necesaria para la interacción con el ARN	97
Figura 28: El complejo THO sh-Tho2 también interacciona con la proteína Tex1	98
Figura 29: El complejo THO sh-Tho2 no interacciona con los ácidos nucleicos	99
Figura 30: Análisis del crecimiento y de la expresión del gen lac-Z en las diferentes cepas de <i>S. cerevisiae</i>	102

DISCUSIÓN

Figura 31: Componentes del complejo TREX de <i>S. cerevisiae</i>	107
Figura 32: Proteínas que interaccionan con el complejo THO de <i>S. cerevisiae</i> durante su función	109
Figura 33: Componentes de los complejos THO y TREX conservados en la evolución	111
Figura 34: Localización de las proteínas Hpr1 y Tex1 en el complejo THO:Tex1	113
Figura 35: Predicción de la estructura secundaria de Tex1 en las regiones amino y carboxi-terminal	115
Figura 36: Composición en Argininas de la proteína Tho2 de <i>S. cerevisiae</i>	116
Figura 37: El extremo carboxi-terminal de Tho2 está aparentemente desestructurado	118
Figura 38: Disposición de la secuencia de la proteína Hpr1 de <i>S. cerevisiae</i>	120
Figura 39: Proteínas que interaccionan con Hpr1 en <i>S. cerevisiae</i>	121
Figura 40: Modelo esquemático de un posible ciclo de expresión génica en <i>S. cerevisiae</i>	129



Abreviaturas

Å: Angstrom
 ADN: Ácido desoxirribonucleico
 AID: Deaminasa inducida por activación (*Activation-Induced Cytidine Deaminase*)
 ARN: Ácido ribonucleico
 ARNm: Ácido ribonucleico mensajero
 ARNP_{II}: ARN polimerasa de tipo II
 ATP: Adenosina trifosfato
 BSOF_T: Paquete informático para tratamiento de imágenes de biología estructural (*Bernard's Software Package*)
 CBC: Complejo de unión a la caperuza 5' (*Cap Binding Complex*)
 CBP: Péptido de unión a calmodulina (*Calmodulin Binding Peptide*)
 CTD: Dominio Carboxi-terminal de la ARNP_{II} (Carboxi-Terminal Domain)
 CTF: Función de transferencia de contraste (*Contrast Transfer Function*)
 DTT: Ditiotreitól
 EDTA: Ácido etilenodiamínico tetraacético
 EJC: Complejo de unión de exones (*Exon Junction Complex*)
 EMAN: Paquete informático de análisis de micrografías (*Electron Micrograph Analysis*)
 FSC: Correlación de los anillos de Fourier (*Fourier Shell Correlation*).
 HEPES: del inglés *N*-2-Hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid
 Hpr1: Proteína hiper-recombinatoria 1 (*Hiperrecombination protein 1*)
 kDa: Kilo Dalton
 YPD: Medio de cultivo de levaduras.
 KerDenSOM: Mapa auto-organizativo basado en la estimación de la densidad de probabilidad (*Kernel Density Estimator Self-Organizing Map*).
 MALDI-TOF: Desorción/Ionización mediante láser asistida por matriz con tiempo de vuelo
 Mft1: Proteína de marcaje de fusión mitocondrial (*Mitochondrial Fusion Targeting 1*)
 NaCl: Cloruro sódico o sal común
 NER: Reparación por escisión de nucleótidos (*Nucleotide Excision Repair*)
 NPC: Complejo de poros de la membrana nuclear (*Nuclear Pore Complex*)
 M: Molar
 mM: mili molar
 ME: Microscopia Electrónica
 ml_align2d: del inglés, *Maximum-likelihood alignment in 2D*
 MwM: Marcadores de peso molecular (*Molecular weight Markers*)

PAGE: Electroforesis en geles de poliacrilamida (*Polyacrylamide Gel Electrophoresis*)

PBS: Tampón fosfato salino

PDB: Banco de datos de proteínas

PMSF: Fluoruro de fenilmetilsulfonilo

PSA: Persulfato amónico

SDS: Dodecil sulfato sódico

SDS-PAGE: Electroforesis desnaturizante en geles de poliacrilamida

SOM: Mapa auto-organizativos (*Self-Organizing Map*)

Sub2: Proteína supresora de la Brr1 (*suppressor of Brr1 protein 2*)

RMN Resonancia magnética nuclear

RNPm: Complejo ribonucleoproteico de ARNm

rpm: revoluciones por minuto

SVD: Análisis estadístico por decomposición singular de valores (*Singular Value Decomposition*)

TAP: Purificación en tándem por afinidad (*Tandem Affinity Purification*).

TCR: Reparación acoplada a la transcripción (*transcription coupled repair*)

TEMED: N,N,N', N' tetra-metil-etilen-diamina

TEV: Proteasa del virus del tabaco (*Tobacco Etch Virus*)

Tho2: Proteína supresora del defecto transcripcional de Hpr1 (*suppressors of the transcriptional defects of hpr1 by overexpression*)

Tris-HCL pH 8,0: Tampón Tris base equilibrado con ácido clorhídrico hasta pH 8,0

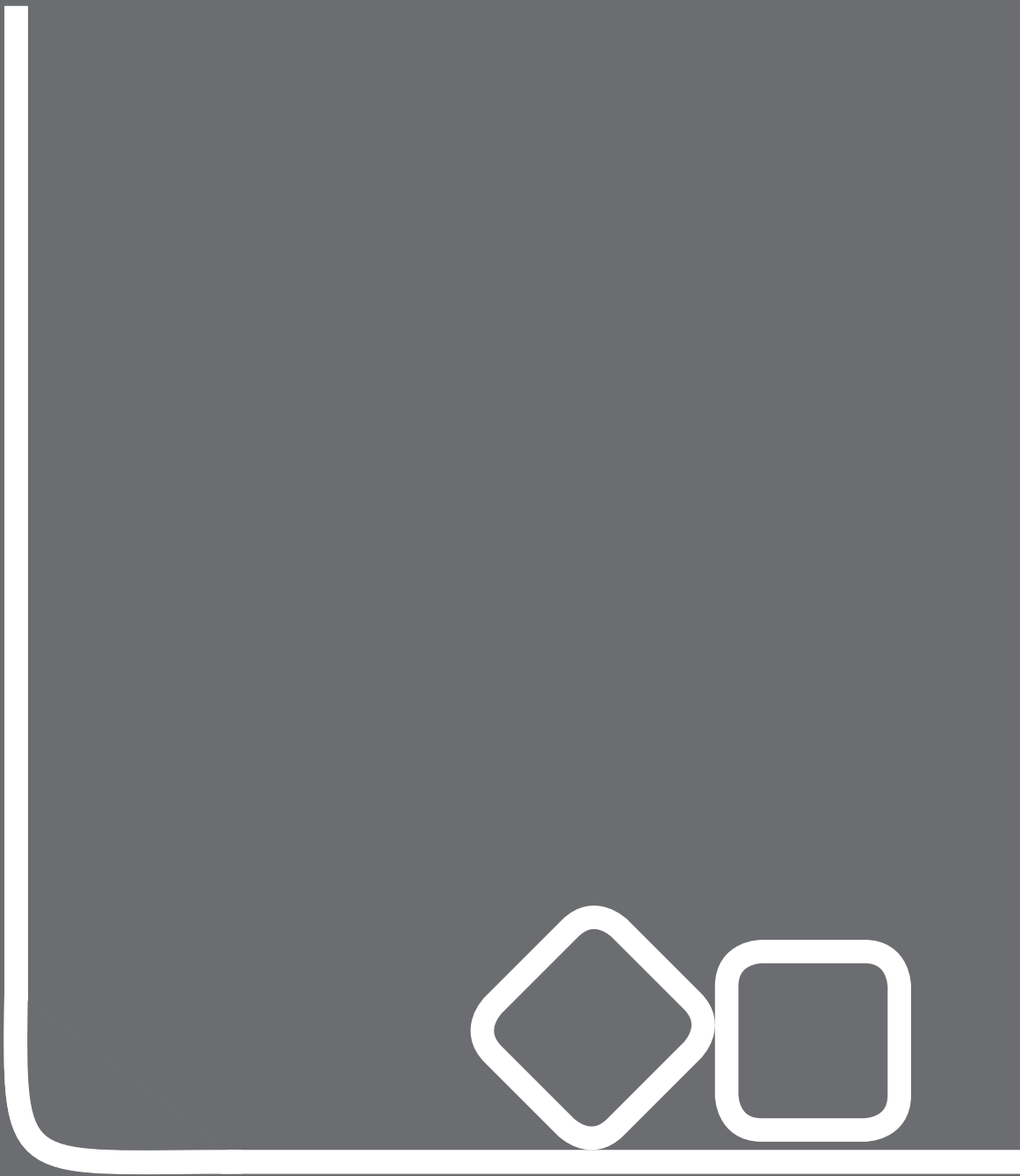
UBA: Dominio asociado a ubiquitina (*ubiquitin associated*)

µg: microgramos

µl: microlitros

µm: micras

XMIPP: Paquete informático de procesamiento y análisis de imagen para microscopia electrónica (*X-Window-based Microscopy Image Processing Package*)





Summary

In eukaryotic cells the genetic information is protected and stored in the nucleus. In order to perform its function, genetic material has to be copied to a temporal ribonucleotide molecule termed messenger RNA (mRNA). The mRNA traverses the nuclear envelope carrying the genetic information to ribosomes, the sites of protein synthesis in the cytoplasm. mRNA's export requires the recruitment of export factors into the mRNA, forming the ribonucleoprotein complexes or mRNPs.

This thesis is focused on the structural characterization of the *Saccharomyces cerevisiae* THO complex, which is directly implicated in the mRNP biogenesis. THO, a four-protein complex (Tho2, Hpr1, Mft1 y Thp2), assists the interaction between the export factors and the mRNA, and facilitates the genetic information flow from the nucleus to the cytoplasm. The THO complex expedites gene expression and at the same time contributes to the cell genetic-stability and collaborates with the quality control mechanisms that ensure the production of active mRNA molecules.

This work has served to generate for the first time a three-dimensional reconstruction of THO complex. Furthermore, it has been indentified the interaction between the THO complex and the Tex1 protein, which interacts with the complex *in vivo* during the transcription process. Moreover, the localization inside the complex of one of the most important THO's components, Hpr1, has been determined. Finally, the specific region of THO complex involved in the interaction with nucleic acids, has been identified and its importance for the cellular survival has been tested. Bringing together these results, it has been suggested a new action-model for the initial steps of the genetic expression in *S. cerevisiae*, that contributes to the understanding of this process in agreement with the current knowledge.





1. Introducción

1.1. Flujo de información y expresión génica en células eucariotas

El ácido desoxirribonucleico (ADN) es la molécula biológica que contiene la información genética que se transmite de unas generaciones a otras. En las células eucariotas esta información se encuentra protegida y almacenada dentro del núcleo celular (Hearst *et al.*, 1970; Macgillivray *et al.*, 1972; Corry *et al.*, 2005). Sin embargo, la información genética cumple su propósito en el citoplasma celular, donde es utilizada para la síntesis de proteínas. Las proteínas son biomoléculas formadas por cadenas lineales de aminoácidos que desempeñan papeles fundamentales en los seres vivos, respondiendo a funciones como el soporte estructural de células y tejidos, la regulación de procesos, el transporte de sustancias, la defensa e identificación de patógenos, la digestión de compuestos o el movimiento. Así, las proteínas son las máquinas moleculares de las células, las biomoléculas mas abundantes de las células y las encargadas de hacer todas aquellas tareas que permiten a la célula desarrollarse en el medio o como parte de un organismo. La composición de proteínas de una célula determina su estructura y su función (Gómez-Moreno *et al.*, 2003), siendo su concentración en el interior celular de entre 200 y 400 miligramos por mililitro, según del tipo de célula (Richter *et al.*, 2008).

La expresión génica incluye todos los procesos necesarios para que la información del ADN llegue a ser utilizada para la producción citoplasmática de proteínas. Estos procesos podrían resumirse en tres etapas fundamentales: la transcripción, la exportación y la traducción (Dever, 2002). La transcripción es un proceso que tiene lugar dentro del núcleo celular, en el cual la información genética del ADN es copiada a una molécula similar de ácido ribonucleoproteico, llamada ARNm (Gómez-Moreno *et al.*, 2003). Una vez finalizado este proceso, el ARNm debe viajar desde el núcleo hasta el citoplasma celular, en un proceso llamado exportación (Rodríguez *et al.*, 2004). Para que la exportación del ARNm sea posible es necesario que esta molécula se asocie con varias proteínas (factores de exportación) formando los llamados complejos ribonucleoproteicos de ARNm (RNPm) que constituyen la forma en la que el ARNm atraviesa la membrana nuclear y viaja hasta el citoplasma celular (Saguez *et al.*, 2005). La formación de los RNPm se conoce como biogénesis

de RNPM (Peng *et al.*, 2003; Luna *et al.*, 2008). Finalmente, en el citoplasma tiene lugar el proceso de traducción, en el cual la secuencia del ARNm es liberada tras el desensamblaje de los RNPM y “leída” por los ribosomas, que dirigen la síntesis de proteínas (Matsumoto, 2006; Wower *et al.*, 2008). Finalizada la traducción, la molécula de ARNm es degradada en el citoplasma, los componentes de los complejos RNPM son importados de nuevo al núcleo y las proteínas recién sintetizadas atraviesan una serie de cambios hasta convertirse en moléculas biológicamente activas (Valpuesta *et al.*, 2002; Hieronymus *et al.*, 2004a; Matsumoto, 2006).

En los últimos años se ha comprobado que los procesos de transcripción, exportación y biogénesis de RNPM están estrechamente relacionados y que existen varios mecanismos que promueven que estos procesos ocurran de forma secuencial, sincronizada y eficaz. El acoplamiento de estos procesos agiliza la expresión génica y maximiza el aprovechamiento de los recursos celulares (Strasser *et al.*, 2002; Reed, 2003; Hieronymus *et al.*, 2004b; Rodríguez *et al.*, 2004; Voynov *et al.*, 2006; Luna *et al.*, 2008). A continuación se explican brevemente los procesos que suceden en las fases iniciales de la expresión génica en células eucariotas.

1.2. Transcripción

La transcripción es un proceso esencial para la supervivencia celular que permite la expresión de la información genética a través de la síntesis de ARN. Se trata del primer paso de la ruta de expresión genética, mediante el cual se transfiere la información contenida en la secuencia de ADN a una secuencia copia, llamada ARN (Gómez-Moreno *et al.*, 2003). Este proceso es dirigido por unas enzimas denominadas ARN polimerasas (ARNP) que sintetizan la molécula de ARN a partir de la secuencia del ADN (Jeffares *et al.*, 1998). La transcripción es por tanto el proceso de síntesis de ARN dirigido por la ARNP (Jukes, 1963; Imamoto, 1973; Nelson, 2000).

En células eucariotas existen diferentes ARN polimerasas que sintetizan distintos tipos de ARN, cada uno de los cuales es exportado al citoplasma a través de una ruta independiente (Nelson, 2000). El ARN mensajero (ARNm), responsable

de la síntesis de proteínas en el citoplasma celular, es sintetizado por la ARN polimerasa II (ARNPII). Para sintetizar el ARNm, la ARNPII se asocia con factores de transcripción y recorre el gen (fragmento de ADN) copiando la información contenida en la secuencia del ADN en la molécula de ARNm (Nelson, 2000; Mathews, 2002). La evolución del proceso de transcripción se puede dividir en tres etapas fundamentales: iniciación, elongación y terminación (Mathews, 2002).

La iniciación de la transcripción es la fase en la que el ADN es preparado para ser copiado. Para ello son necesarias una serie de proteínas (factores de iniciación) que se unen a secuencias específicas de ADN (secuencias promotor) indicando el inicio del gen y definiendo el lugar donde la transcripción ha de comenzar (en el extremo 5' del gen). Junto a los factores de iniciación se une también la ARNPII, que se sitúa entre las dos hebras de ADN interaccionando con una de ellas (hebra molde). La ARNPII comienza la síntesis de la molécula de ARNm uniendo ribonucleótidos por complementariedad de bases con la secuencia de ADN (Voorma *et al.*, 1994; Mathews, 2002). Así comienza la elongación transcripcional, fase en la que la ARNPII recorre el gen catalizando la formación de puentes de hidrógeno entre la hebra molde de ADN y la cadena de ARNm que está siendo sintetizada, al tiempo que promueve la formación de enlaces fosfodiéster entre los ribonucleótidos (Mathews, 2002). Por último, la terminación de la transcripción ocurre cuando la ARNPII reconoce el final de gen que viene marcado por una secuencia específica en el ADN o por unas proteínas que interaccionan específicamente con éste (factores de terminación). La ARNPII se detiene y se desprende del gen, liberando así la molécula de ARNm recién sintetizado.

En los últimos años se ha visto que el proceso de transcripción solapa temporal y espacialmente con los procesos de maduración, biogénesis y exportación citoplasmática de RNPm, de forma que la molécula de ARNm recién sintetizada va siendo simultáneamente procesada por otras proteínas, al tiempo que es sintetizada (Reed, 2003; Aguilera, 2005). Así, cuando termina la transcripción y se libera el ARNm gran parte de la molécula ya ha sufrido un proceso de maduración y empaquetamiento que la hace apta para el transporte núcleo-citoplasmático, donde servirá de molde para la síntesis de nuevas proteínas (Nelson, 2000; Mathews, 2002).

1.3. Maduración del ARN mensajero

El ARNm saliente de la transcripción se conoce como transcrito primario o pre-ARNm. En la mayoría de los casos, esta molécula recién sintetizada debe sufrir una serie de modificaciones antes de ser transportada hacia el citoplasma para ejercer allí su función (Mathews, 2002). Estas modificaciones son lo que se conoce como maduración del ARNm y ocurren al mismo tiempo que la molécula está siendo sintetizada, es decir, cotranscripcionalmente (Kataoka *et al.*, 2000; Afasizhev, 2007; Lin *et al.*, 2008).

El primer paso en la maduración del pre-ARNm ocurre al inicio de la transcripción, cuando se añade una estructura denominada caperuza 5' en el comienzo de la molécula (extremo 5') (Sims *et al.*, 2004; Cole *et al.*, 2009) (Figura 1). La caperuza 5' consta de un nucleótido modificado de guanidina que interacciona con el fosfato libre de la posición 5' de la cadena de ARNm. La formación de la caperuza 5' se lleva a cabo por un complejo de proteínas especializadas llamado CBC (*Cap Binding Complex*) cuya acción es regulada por la ARNP II (Cowling, 2010). La adición de la caperuza 5' en el comienzo de la molécula es necesaria para el desarrollo de la transcripción y para su exportación citoplasmática. Además, la caperuza 5' protege al ARNm de la degradación (Grudzien *et al.*, 2006) y actúa en el reconocimiento y acceso al ribosoma (Dever, 1999; Stachelska *et al.*, 2002; Cheng *et al.*, 2006). La eliminación de la caperuza 5' promueve la degradación citoplasmática del ARNm una vez terminada su función (Parker *et al.*, 2007).

En células eucariotas la mayoría de los genes tienen secuencias de ADN no codificantes, llamadas intrones, que no aportan información útil para la síntesis de proteínas. Durante la transcripción los intrones se transcriben con el resto del gen y pasan a formar parte de la molécula de pre-ARNm. Para que la molécula pueda ejercer su función en el citoplasma los intrones deben ser eliminados del pre-ARNm (Flint, 1981; Cech, 1983). Esto ocurre por medio de un proceso denominado ajuste o *splicing*, que consiste en el corte y empalme de la secuencia del pre-ARNm, de forma que se eliminen los intrones dando como resultado una molécula de ARNm en la que toda la secuencia contiene información para la síntesis de proteínas. El

proceso de *splicing* comienza con el reclutamiento cotranscripcional de proteínas que interaccionan con la molécula de pre-ARNm y favorecen la formación del spliceosoma (Lamond, 1993; Wise, 1993; Kramer, 1996). El spliceosoma es un complejo proteico de gran tamaño que se encarga de cortar, descartar los intrones y unir los fragmentos codificantes de la secuencia del ARNm para obtener la molécula de ARNm lista para ejercer su función en el citoplasma (Newman, 1997; Tarn *et al.*, 1997). El spliceosoma colabora también con las proteínas del llamado complejo EJC (*Exon Junction Complex*) que se encargan de marcar el principio y el final de los intrones, interaccionando específicamente con la molécula de ARNm (Custodio *et al.*, 2004) (Figura 1). Además, existe un fenómeno llamado “*splicing* alternativo” que permite ajustar el tamaño y la composición de las secuencias del ARNm maduro, de forma que pueden obtenerse moléculas diferentes de ARNm a partir de un mismo gen. Este hecho multiplica la cantidad de posibles proteínas que pueden ser producidas por cada gen y potencia la variabilidad de biomoléculas producidas por cada célula (Ast, 2004).

El último paso en la maduración del ARNm ocurre cuando la ARNP_{II} reconoce el final del gen. En este punto tiene lugar la poliadenilación del extremo 3' de la molécula recién sintetizada de ARNm, que consiste en un corte endonucleotídico seguido de la adición de una secuencia de bases adenina al final de la molécula (extremo 3'), conocida como cola de poliadenilación 3' (Sheets *et al.*, 1989; Proudfoot, 2004). La adición de esta cola de poliadenilación 3' es fundamental para el transcurso de la transcripción, protege al ARNm de la degradación (Butler, 2002; Grudzien *et al.*, 2006) y favorece su exportación (Jarmolowski *et al.*, 1994; Huang *et al.*, 1996; Hilleren *et al.*, 2001). La cola de poliadenilación 3' también participa en la localización subcelular del ARNm (Hilleren *et al.*, 2001; Jensen *et al.*, 2001b; Jensen *et al.*, 2001a; Libri *et al.*, 2002) y favorece su liberación del lugar de la transcripción (Libri *et al.*, 2002; Thomsen *et al.*, 2003) (Figura 1).

Una vez que se ha añadido la caperuza 5', se han eliminado los intrones y se ha formado la cola de poliadenilación 3', se obtiene una molécula activa conocida como ARNm maduro (Libri *et al.*, 2002; Mathews, 2002). Esta molécula debe ser transportada al citoplasma para dirigir allí la síntesis proteica. Para ello, el ARNm

maduro se asocia con otras proteínas (factores de exportación) dando lugar a unas estructuras compactas llamadas RNPm (Saguez *et al.*, 2005) (Figura 1). El proceso de formación de RNPm se conoce como biogénesis de RNPm y ocurre al mismo tiempo que se sintetiza y procesa el ARNm (Peng *et al.*, 2003; Olesen *et al.*, 2005).

1.4. Biogénesis de RNPm

Para ser transportado desde el núcleo hasta el citoplasma celular, el ARNm maduro ha de ser recubierto con una serie de proteínas que favorecen su translocación a través de la membrana nuclear, lo dirigen hacia su destino en el citoplasma y lo protegen de la degradación y de las interacciones no deseadas con otras macromoléculas presentes en el citoplasma. Este proceso recibe el nombre de biogénesis de RNPm (Olesen *et al.*, 2005; Saguez *et al.*, 2005).

El complejo RNPm es la estructura que adopta el ARNm para “viajar” desde el núcleo hasta el citoplasma celular. Los componentes del RNPm interaccionan entre sí y se organizan en una estructura compacta diseñada para ser transportada a través de la membrana nuclear, haciendo la función de “maleta” de la información genética. La biogénesis de el RNPm es imprescindible para la exportación citoplasmática y para la supervivencia citoplasmática del ARNm mensajero en todas las células eucariotas (Spohr *et al.*, 1970; Jimeno *et al.*, 2010). El transporte núcleo-citoplasmático está regulado por los receptores de la membrana nuclear. Así, la exportación de RNPm tiene lugar gracias a interacciones específicas entre sus componentes y el complejo de poros de la membrana nuclear (*Nuclear Pore Complex* o NPC). El RNPm se ajusta en el interior del canal central del NPC y fluye a través de éste hasta alcanzar el citoplasma celular (Zenklusen *et al.*, 2001; Custodio *et al.*, 2004; Erkmann *et al.*, 2004; Stewart, 2007) (Figura 1).

En los últimos años se ha descubierto que los factores responsables de la biogénesis de RNPm son reclutados cotranscripcionalmente (Chávez *et al.*, 2000; Strasser *et al.*, 2002). Este hecho supone la existencia de una conexión funcional entre las etapas iniciales de la expresión génica. Así, los procesos de transcripción, maduración, biogénesis y exportación de RNPm están acoplados y se suceden de forma

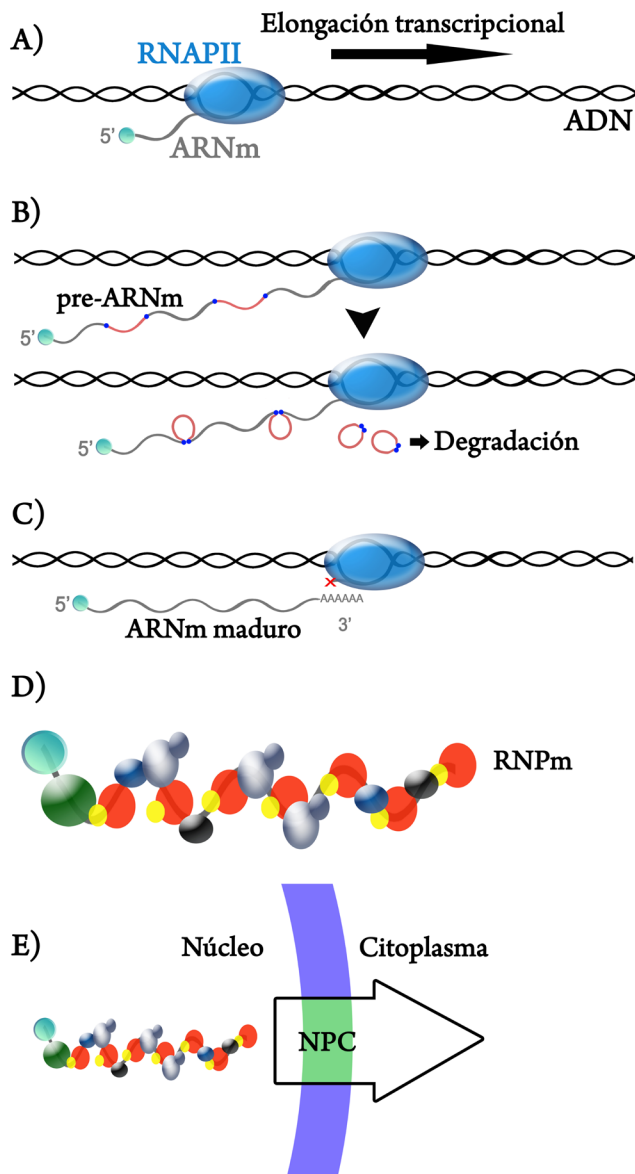


Figura 1: Etapas iniciales de la expresión génica en células eucariotas. **A)** Adición de la caperuza 5'. Al comienzo de la transcripción se añade un complejo proteico en el inicio de la molécula de ARNm (extremo 5') que se denomina caperuza 5' (representada como una circunferencia verde). **B)** *Splicing*. A medida que avanza la transcripción, el ARNm recién sintetizado (pre-ARNm) es sometido a un proceso de corte y empalme de secuencias conocido como *splicing*, en el que se descartan los intrones (representados como segmentos de color rojo). Los extremos de los intrones son marcados por el complejo EJC que interaccionan específicamente con el ARNm (representado en azul oscuro). **C)** Poliadenilación 3'. La adición de la cola de poliadenilación en el extremo final de la molécula de ARNm (extremo 3') favorece el corte de la de ésta en su extremo final y la correspondiente liberación del ARNm de la RNAPII. El ARNm resultante de estos tres pasos se conoce como ARNm maduro. **D)** Biogénesis de RNPm. Para ser exportado al citoplasma el ARNm maduro es recubierto por proteínas que lo protegen y dirigen hacia el citoplasma celular. La estructura compacta formada por el ARNm y las proteínas que lo recubren se llama complejo ribonucleoproteico de ARNm (RNPm). **E)** Exportación. El RNPm atraviesa la membrana nuclear (en azul) a través del complejo de poros de la membrana (NPC, en verde) y viaja al citoplasma para ejercer allí su función: la síntesis proteica. Para simplificar el esquema, se han representado las primeras fases de la expresión génica como procesos independientes aunque en la célula, estos procesos están acoplados y coinciden en espacio y tiempo unos con otros (Reed, 2003).

sincronizada y secuencial coincidiendo en espacio y tiempo unos con otros (Rondon *et al.*, 2003b; Saguez *et al.*, 2008). El acoplamiento de estos procesos agiliza la evolución de todos ellos, facilita el aprovechamiento de los recursos celulares y maximiza la eficacia de la expresión génica en todos los organismos eucariotas (Hammell *et al.*, 2002; Reed, 2003; Rondon *et al.*, 2003b; Hieronymus *et al.*, 2004b). El dominio carboxi-terminal de la ARNP_{II}, conocido como CTD (*Carboxi-Terminal Domain*) ejerce una función importante en la regulación y coordinación de estos procesos (Strasser *et al.*, 2002). Durante la elongación transcripcional, el ARNm emerge de la ARNP_{II} por la base del CTD de forma que éste actúa como “plataforma” para los factores de maduración y biogénesis de RNP_m (Jimeno *et al.*, 2008). El CTD presenta diversos estados de fosforilación que regulan el reclutamiento secuencial de factores a lo largo de la elongación transcripcional (Masuda *et al.*, 2005). El orden en el que son incorporados los diferentes factores es aún desconocido pero se sabe que en la cadena de reclutamiento de factores regulada por el CTD, algunos factores ayudan a la incorporación de otros (Olesen *et al.*, 2005).

La biogénesis de RNP_m es también un mecanismo de control de calidad del proceso de producción de ARNm, en el que solo son empaquetados en RNP_m y exportados al citoplasma aquellos ARNm que han sido producidos correctamente y están definitivamente listos para ser traducidos a proteínas en el citoplasma celular (Stutz *et al.*, 2003; Saguez *et al.*, 2005). La cadena de reclutamiento secuencial de factores de exportación sobre el ARNm requiere que los factores sean correctamente incorporados sobre el ARNm, para permitir su exportación citoplasmática (Hurt *et al.*, 2000; Strasser *et al.*, 2001). El hecho de que unos factores sean responsables del reclutamiento de otros supone que si la incorporación de alguno de ellos falla, puede fallar también la incorporación de otros. Errores en la biogénesis de RNP_m pueden inducir la degradación del ARNm de forma que si uno de ellos no es correctamente reclutado sobre el ARNm puede detenerse la transcripción e incluso derivar el ARNm hacia su degradación (Hilleren *et al.*, 2001; Libri *et al.*, 2002; Olesen *et al.*, 2005). Estos mecanismos de control son de gran importancia para la célula dado que la exportación de ARNm defectuosos podría suponer la producción de proteínas inservibles y potencialmente perjudiciales para la célula (Kubota, 2009).

En organismos eucariotas existe un complejo multiproteico que esta directamente implicado en el acoplamiento de las etapas iniciales de la expresión génica (Strasser *et al.*, 2002; Rondon *et al.*, 2003b; Luna *et al.*, 2008; Jimeno *et al.*, 2010). Este complejo recibe el nombre de “complejo THO” y debe su nombre al componente mayoritario del complejo, la proteína Tho2 (*suppressors of the transcriptional defects of hpr1 by overexpression*). El complejo THO dirige la biogénesis de RNPM y coordina los procesos de transcripción y exportación de ARNm a la vez que mantiene la estabilidad genética y contribuye a la supervivencia celular en condiciones desfavorables (Luo *et al.*, 1999; Jensen *et al.*, 2001b; Cheng *et al.*, 2006).

1.5. El complejo THO: perspectiva

Desde hace mas de cuarenta años se sabe que el ARNm de eucariotas es exportado al citoplasma celular en forma de complejos RNPM (Spirin *et al.*, 1965; Spohr *et al.*, 1970; Kumar *et al.*, 1972). Sin embargo, el mecanismo por el cual se forman estas estructuras compactas de ARNm y proteínas no se conoce aún con exactitud (Jimeno *et al.*, 2010). Durante los últimos años numerosos grupos de investigación han ido aportando resultados que contribuyen al avance en la comprensión de este complicado proceso que tiene lugar de forma continua en el núcleo de todas las células eucariotas, asociado a la transcripción. Para entender el proceso de biogénesis de RNPM y el acoplamiento de los procesos de transcripción, maduración y exportación de ARNm, ha tenido especial relevancia el descubrimiento del complejo THO y los factores que interaccionan con él. A continuación se exponen en orden cronológico algunos de los descubrimientos mas significativos que han conducido hasta el conocimiento actual (Figura 2).

En el año 1973 se descubrió que la formación de RNPM es un hecho imprescindible para la exportación del ARNm en todas las células eucariotas (Wall, 1973). A finales de los años ochenta se identificó la proteína Hpr1 (*Hiperrecombination protein 1*) de *Saccharomyces cerevisiae*, el primer componente del complejo THO en ser descrito (Aguilera *et al.*, 1988; 1989; 1990). Hpr1 se identificó por su capacidad de generar inestabilidad genética durante la transcripción pero por entonces

no se conocía ninguna de las proteínas implicadas en el mecanismo de biogénesis de RNPm y no se la relacionó con este proceso. Poco después se identificó la proteína Mft1 (*Mitochondrial Fusion Targeting 1*) que se relacionó inicialmente con los mecanismos de importación proteica y marcaje mitocondrial en lavaduras (Garrett *et al.*, 1991). Con el tiempo se observó que tanto Hpr1 como Mft1 intervienen en las fases tempranas del proceso de expresión génica (Santos-Rosa *et al.*, 1994) y se identificó la proteína homóloga de Hpr1 en humanos (Thoc1), lo que llevó a proponer que su función está conservada a lo largo de la evolución desde levaduras hasta humanos (Durfée *et al.*, 1994).

El primer indicio de la existencia de un complejo implicado directamente en la biogénesis de RNPm no llegó hasta el año 1995, con el descubrimiento de que la proteína Hpr1 de *S. cerevisiae* forma parte de un complejo de gran tamaño implicado en la regulación positiva de la transcripción, que promueve la síntesis de ARNm agilizando la elongación transcripcional (Zhu *et al.*, 1995; Piruat *et al.*, 1996; Chávez *et al.*, 1997). En 1998 se descubrió la proteína Tho2 (*suppressors of the transcriptional defects of hpr1 by overexpression 2*) de *S. cerevisiae*, el tercer y mayor componente del complejo THO, cuya función se relacionó inmediatamente con la de Hpr1 (Piruat *et al.*, 1998). En 1999 se observó que Hpr1 interacciona directamente con la ARNP_{II} y con la maquinaria de poliadenilación del extremo 3' del ARNm (Chang *et al.*, 1999; Schneiter *et al.*, 1999). Estos descubrimientos llevaron a la conclusión de que existe un acoplamiento entre los procesos de síntesis, maduración, biogénesis y exportación de ARNm que dirige los transcritos hacia el citoplasma y agiliza la expresión génica en células eucariotas (Luo *et al.*, 1999; Kataoka *et al.*, 2000). Finalmente, en el año 2000 se describió el complejo THO de *S. cerevisiae* y con él se identificó el cuarto y último componente del complejo, la proteína Thp2 (Chávez *et al.*, 2000). El complejo THO se relacionó desde el primer momento con la biogénesis de RNPm (Chávez *et al.*, 2001; Jensen *et al.*, 2001b) y pronto se propuso su implicación en los mecanismos de control de calidad del ARNm (Hilleren *et al.*, 2001; Jensen *et al.*, 2001a).

El descubrimiento del complejo THO supuso un avance importante en la comprensión del proceso de biogénesis de RNPm y del acoplamiento de los

procesos de la expresión génica. Sin embargo, el modelo de acción del complejo fue propuesto posteriormente gracias a la caracterización de las interacciones entre el complejo THO y otros factores de transcripción y exportación. En el año 2001 se observó que el complejo THO interacciona directamente con el factor de exportación Mex67 (Strasser *et al.*, 2001; Zenklusen *et al.*, 2001) y en el año 2002 se identificaron varias proteínas que interaccionan de forma estable con el complejo THO, contribuyendo a su función (Strasser *et al.*, 2002). Se observó también que el complejo THO, junto con otros factores de exportación, se asocia directamente con los ácidos nucleicos ADN y ARN durante la transcripción y que interviene en los mecanismos de reparación del ADN (Jimeno *et al.*, 2002; Smolka *et al.*, 2007), en el *splicing* de metazoos (Rappsilber *et al.*, 2002; Zhou *et al.*, 2002; Jurica *et al.*, 2003) y en la degradación de transcritos defectuosos, a través de su interacción con el exosoma nuclear (Libri *et al.*, 2002). Se consideró al complejo THO como responsable de la formación de RNPM y así, se propuso que su actividad determina el destino y la expresión de los ARNm en el citoplasma celular (Hieronymus *et al.*, 2003; Keene, 2003). En el año 2003 se observó que el complejo THO previene la formación de híbridos ADN:ARN (llamados bucles-R o R-loops) y contribuye a la abolición del daño genético y la supervivencia celular (Huertas *et al.*, 2003). Se identificaron los factores Hrb1 y Gbp2 asociados de forma estable con el complejo THO (Windgassen *et al.*, 2003; Hurt *et al.*, 2004). En el año 2004, se describió el complejo THO de *Drosophila melanogaster* (Rehwinkel *et al.*, 2004) y se observó que la función del complejo THO es especialmente importante para la supervivencia celular en condiciones de estrés celular (Hacker *et al.*, 2004; Hieronymus *et al.*, 2004b; Rehwinkel *et al.*, 2004).

En los últimos años la demostración de que la proteína Hpr1 es susceptible de ubiquitinación ha llevado a proponer a esta proteína como la responsable de la estabilidad del complejo THO (Gwizdek *et al.*, 2005). Además, Hpr1 se ha visto directamente relacionada con el desarrollo del cáncer de mama (Guo *et al.*, 2005) y se ha descrito que su interacción con el factor de exportación Mex67 previene la degradación del complejo THO (Gwizdek *et al.*, 2006). En *S. cerevisiae* se ha observado que las proteínas Tho2 y Hpr1 son los componentes más relevantes para la

- 1990 Identificación de la proteína Hpr1 en *Saccharomyces cerevisiae*
- 1991 Identificación de la proteína Mft1 en *S. cerevisiae*
- 1994 Identificación de la proteína homóloga de Hpr1 en humanos (Thoc1)
- 1995 Hpr1 regula positivamente el proceso de transcripción en *S. cerevisiae*
Hpr1 forma parte de un complejo de gran tamaño en *S. cerevisiae*
- 1996 Hpr1 promueve la elongación transcripcional en *S. cerevisiae*
- 1998 Identificación de la proteína Tho2 en *S. cerevisiae*
- Hpr1 está relacionada con la maquinaria de poliadenilación del extremo 3' del ARNm en *S. cerevisiae*
- 1999 Hpr1 interactúa directamente con la ARN polimerasa II de *S. cerevisiae*
Los procesos de transcripción, maduración, biogénesis y exportación de ARNm están acoplados en eucariotas
- 2000 Identificación del complejo THO y de la proteína Thp2 de *S. cerevisiae*
- THO está implicado en la biogénesis de RNPm en *S. cerevisiae*
- 2001 THO interviene en los mecanismos de control de calidad de ARNm en *S. cerevisiae*
THO interactúa con Mex67 en *S. cerevisiae*
- THO interactúa de forma estable con las proteínas Tex1, Yra1 y Sub2 en humanos y en *S. cerevisiae*
El complejo THO recorre los genes interactuando con la ARNPm en *S. cerevisiae*
- 2002 El complejo THO, junto con otros factores, interactúa con el ADN y con el ARN en *S. cerevisiae*
El complejo THO interviene en la reparación cotranscripcional de ADN y previene el daño genético en *S. cerevisiae*
El complejo THO interactúa con el exosoma y regula los mecanismos de control de calidad de ARNm en *S. cerevisiae*
El complejo THO está interactuando con la maquinaria de *splicing* en metazoos
- El complejo THO previene la formación de híbridos ADN:ARN en *S. cerevisiae*
- 2003 El complejo THO determina la composición de los RNPm, el destino y la expresión de los ARNm en *S. cerevisiae*
El complejo THO recluta las proteínas Hrb1 y Gbp2 sobre el ARNm cotranscripcionalmente en *S. cerevisiae*
- Identificación del complejo THO de *Drosophila melanogaster*
- 2004 La función del complejo THO es especialmente importante en condiciones de estrés celular en *D. melanogaster*
El complejo THO se disocia de la ARNPm después de la poliadenilación del extremo 3' en *S. cerevisiae*
- 2005 Hpr1 es susceptible de ubiquitinación y controla la estabilidad del complejo THO en *S. cerevisiae*
El complejo THO de humanos está directamente relacionado con la proliferación del cáncer de mama
- El complejo THO funciona en el mantenimiento cotranscripcional de la topología del ADN en *S. cerevisiae*
- 2006 El factor de exportación Mex67 protege a Hpr1 de la degradación durante la transcripción en *S. cerevisiae*
El complejo THO de humanos es reclutado en el extremo 5' del ARNm al inicio de la transcripción
- 2007 El complejo THO interviene en la reparación cotranscripcional del ADN y en la estabilidad genómica
- Tho2 y Hpr1 son los componentes más relevantes para el desarrollo de la función fisiológica del complejo THO
- 2008 El complejo THO colabora en la poliadenilación del extremo 3' del ARNm en *S. cerevisiae*
La función del complejo THO se hace más necesaria cuanto mayor es la tasa de expresión del gen en *S. cerevisiae*
- 2009 Hpr1 transfiere Mex67 al NPC al final de la transcripción y desestabiliza el complejo THO de *S. cerevisiae*
- 2010 Los componentes del complejo THO pueden adoptar distintas funciones en los diferentes tejidos celulares de metazoos

Figura 2: **Principales aportaciones en relación al complejo THO en los últimos veinte años.** La figura muestra en orden cronológico algunos de los hechos más destacables que han conducido a la comunidad científica hasta la visión actual del complejo THO y de los procesos de biogénesis y exportación de RNPm en células eucariotas. A la izquierda se ha representado una línea temporal en la que se han resaltado los años en los que han tenido lugar las sucesivas aportaciones. Se observa que desde el año 2000 en el que se descubre la existencia del complejo THO han sido muchos los avances y propuestas que han permitido alcanzar el conocimiento actual. Desde entonces, cada año crece el número de grupos de investigación interesados en estudio del complejo THO y de los mecanismos de biogénesis y exportación de ARNm.

función fisiológica del complejo THO (García-Rubio *et al.*, 2008) y que el complejo contribuye a la formación de la cola de poliadenilación del extremo 3' de ARNm durante la terminación transcripcional (Saguez *et al.*, 2008).

Recientemente, estudios realizados en metazoos han llevado a proponer al complejo THO como una posible diana para tratamientos anti-cancerígenos (Guo *et al.*, 2005) y se ha visto que su función es esencial para la supervivencia de embriones de ratón (Wang *et al.*, 2006b) y para el desarrollo de la médula ósea y la hematopoyesis en mamíferos (Mancini *et al.*, 2010). Actualmente se acepta que en metazoos las distintas subunidades del complejo THO pueden desempeñar funciones distintas en los diferentes tipos celulares (Jimeno *et al.*, 2010) (Figura 2).

1.6. Componentes del complejo THO de *Saccharomyces cerevisiae*

El complejo THO de *S. cerevisiae* es un heterotetrámero nuclear estable compuesto por las proteínas Tho2, Hpr1, Mft1 y Thp2 (Chávez *et al.*, 2000) (Figura 3). Las cuatro subunidades del complejo interaccionan entre sí (Strasser *et al.*, 2002) y el complejo se mantiene estable *in vitro* incluso en condiciones de alta salinidad (Strasser *et al.*, 2002). A pesar de que el complejo THO interacciona también con otras proteínas (factores de maduración y exportación de ARNm), tanto su ensamblaje como su integridad dependen exclusivamente de sus cuatro componentes (Huertas *et al.*, 2006).

La proteína mas grande del complejo THO se denomina Tho2 o Rlr1 (1597 aminoácidos, 180 kDa) y representa aproximadamente la mitad del tamaño total del complejo (Figura 3). Tho2 se identificó inicialmente en el núcleo de *S. cerevisiae* funcionando en las primeras fases del proceso de transcripción junto con la ARN-Pol II (Piruat *et al.*, 1998). Mas adelante se describió su implicación funcional con la proteína Hpr1 (2000; Chávez *et al.*, 2001; Aguilera, 2005) y finalmente se identificó como el componente mas importante para la ejecución de la función fisiológica del complejo THO (García-Rubio *et al.*, 2008). Sorprendentemente, la proteína Tho2 presenta características comunes con algunas chaperonas y con algunas subunidades del proteosoma, como la repetición de motivos KEKE (Lisina-Glutámico), relacio-

nados con la asociación de proteínas (Realini *et al.*, 1994; West *et al.*, 2000). Tho2 es la única proteína del complejo THO que ha sido descrita en plantas, donde parece desempeñar una función en el metabolismo, junto con los microtúbulos, tanto en el núcleo como en el citoplasma celular (Hamada *et al.*, 2009). Tho2 está muy conservada a lo largo de la evolución y han sido descritas proteínas homólogas en los complejos THO de *D. melanogaster* y de *H. sapiens* (Thoc2) (West *et al.*, 2000; Strasser *et al.*, 2002; Reed *et al.*, 2005).

El segundo componente en tamaño del complejo THO se denomina Hpr1 (752 aminoácidos, 90 kDa) (Figura 3) y fue la primera proteína del complejo en ser descrita. Hpr1 fue identificada inicialmente por su capacidad de generar fenotipos hiper-recombinantes en *S. cerevisiae* (Aguilera *et al.*, 1990). Más adelante fue descrita como componente de un complejo que interacciona directamente con la ARN-Pol II (Chang *et al.*, 1999) y durante un tiempo se la consideró parte de la maquinaria de transcripción (Chávez *et al.*, 2001; Fan *et al.*, 2001). Con el descubrimiento del complejo THO, Hpr1 se relacionó directamente con la biogénesis y exportación de RNPm (Jimeno *et al.*, 2002; Libri *et al.*, 2002) y se la atribuyó la función de reclutar el factor de exportación Sub2 (*suppressor of Brr1 protein 2*) sobre el ARNm durante el proceso de transcripción (Libri *et al.*, 2002; Zenklusen *et al.*, 2002). Recientemente se ha visto que Hpr1 es susceptible de ubiquitinación en su extremo carboxi-terminal y se la ha propuesto como responsable del ensamblaje y degradación del complejo THO. Hpr1 interacciona directamente con el factor de exportación Mex67

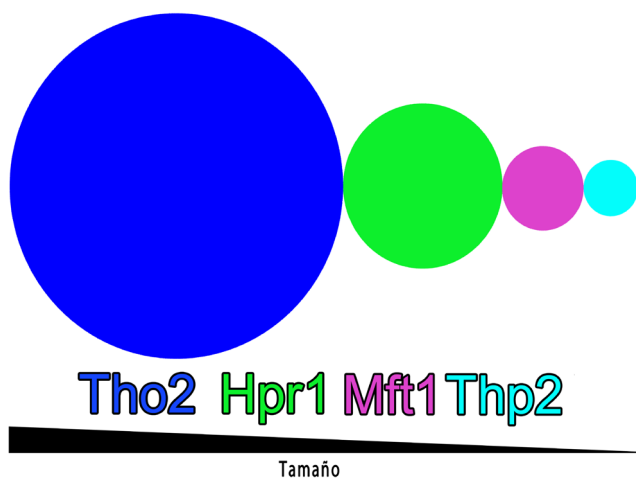


Figura 3: Componentes del complejo THO. El esquema representa los cuatro componentes del complejo THO de *Saccharomyces cerevisiae*, por orden de tamaño. En la figura se aprecia que la proteína de mayor tamaño, Tho2 (azul oscuro) representa aproximadamente la mitad del tamaño total del complejo. La sigue Hpr1 (verde), Mft1 (morado) y Thp2 (azul claro). Los tamaños reales de los componentes del complejo THO son: Tho2 (183 kDa), Hpr1 (87,8 kDa), Mft1 (44,9 kDa) y Thp2 (30,3 kDa). En la imagen, cada proteína está representada por una circunferencia cuyo área es proporcional al tamaño real de la proteína.

que además la protege de la degradación (Gwizdek *et al.*, 2006). Hpr1 ha sido identificada como la segunda proteína más importante para la ejecución de la función fisiológica del complejo THO y su extremo amino-terminal parece ser especialmente importante para su función (García-Rubio *et al.*, 2008). Curiosamente, el aminoácido 101 de Hpr1 puede modificarse provocando una alteración en la biogénesis de RNPm sin afectar a la inestabilidad genética (Huertas *et al.*, 2006; Wellinger *et al.*, 2006). Hpr1 está muy conservada a lo largo de la evolución y ha sido descrita como componente del complejo THO de *D. melanogaster* y de *H. sapiens* (Thoc1) (Strasser *et al.*, 2002; Reed *et al.*, 2005).

La proteína Mft1 (también llamada Mft52) es el siguiente componente del complejo THO en tamaño (392 aminoácidos, 45 kDa) (Figura 3). Mft1 se relacionó en un principio con los mecanismos de importación mitocondrial (Garrett *et al.*, 1991) y se describió como una proteína citoplasmática de carácter fuertemente básico con dos dominios claramente diferenciados. Mas adelante Mft1 se relacionó con las fases finales de la expresión génica y se vio asociada con el ribosoma en el marcaje específico de proteínas mitocondriales (Beilharz *et al.*, 1997; Cartwright *et al.*, 1997). Finalmente se describió como componente del complejo THO en el interior del núcleo (Chávez *et al.*, 2000). Recientemente se ha propuesto su función en el complejo THO como responsable del reclutamiento de los factores de exportación Gbp2 y Hrb1 sobre el ARNm (Windgassen *et al.*, 2003; Hacker *et al.*, 2004; Hurt *et al.*, 2004). Mft1 se asocia también con la regulación del proceso de transcripción dado que interacciona con la kinasa Ctk1, encargada de fosforilar el CTD de la ARNP_{II} durante la elongación transcripcional (Hurt *et al.*, 2004). Mft1 no está conservada en la evolución y sólo ha sido descrita en levaduras (Reed *et al.*, 2005).

Por último, la proteína mas pequeña del complejo THO se denomina Thp2 (261 aminoácidos, 30 kDa) (Figura 3). Thp2 no se conocía antes de ser identificada como componente del complejo THO y es aún la proteína mas desconocida del complejo. Thp2 no está conservada a lo largo de la evolución y sólo ha sido descrita en levaduras (Reed *et al.*, 2005).

1.7. El complejo THO en la expresión génica

Hasta la fecha el complejo THO ha sido identificado en *Saccharomyces cerevisiae* (Chávez *et al.*, 2000), en *Homo sapiens* (Strasser *et al.*, 2002) y en *Drosophila melanogaster* (Rehwinkel *et al.*, 2004) y a pesar de que existen diferencias en algunos de sus componentes, se cree que su función se encuentra muy conservada a lo largo de la evolución (Reed *et al.*, 2002). En todos los organismos, el complejo THO funciona durante la primera fase de la expresión génica contribuyendo a la biogénesis de RNPM, al acoplamiento de los procesos de expresión génica y al mantenimiento de la integridad genética pero su modelo de acción no se conoce aún con exactitud. A pesar de que la función del complejo se encuentra conservada, se han propuesto algunas diferencias en el modelo de acción del complejo en levaduras y en metazoos (Jimeno *et al.*, 2010), que son consistentes con el hecho de que el proceso de transcripción difiere considerablemente entre ambos organismos. Así, en metazoos la cantidad de intrones es mucho mayor que levaduras lo que provoca que la maquinaria de *splicing* esté mucho más desarrollada. Curiosamente, en levaduras el complejo THO parece ser reclutado por la maquinaria de transcripción mientras que en metazoos esta función parece responsabilidad de la maquinaria de *splicing* (Reed *et al.*, 2005). De esta forma se concluye que en levaduras, donde la mayoría de los genes carecen de intrones, el complejo THO funciona ligado a la transcripción mientras que en metazoos, donde la gran mayoría de transcritos presentan intrones, el complejo THO funciona mayoritariamente ligado al proceso de *splicing* (Cheng *et al.*, 2006).

En el modelo de acción aceptado actualmente para *S. cerevisiae*, el complejo THO interviene inicialmente en el ensamblaje del complejo de transcripción y se incorpora a la maquinaria de transcripción a través de la interacción de Hpr1 con la ARNP_{II} (Chang *et al.*, 1999). El complejo THO recorre el gen guiado por la ARNP_{II}, que se encarga de sintetizar el ARNm al tiempo que avanza por la secuencia de ADN (Strasser *et al.*, 2002). El complejo THO facilita la interacción de algunos factores de exportación con el ARNm recién sintetizado, protege a la información genética durante el proceso de transcripción (Gómez-González *et al.*, 2007) y evita

la formación de híbridos de ADN:ARN entre la cadena de ARNm recién sintetizada y la hebra libre de ADN (Gómez-González *et al.*, 2009a). Así, el complejo THO mantiene la estabilidad genética y minimiza la recombinación mitótica asociada a la transcripción (Jimeno *et al.*, 2002; Huertas *et al.*, 2003; Rehwinkel *et al.*, 2004; Wellinger *et al.*, 2006). Cuando la ARNP_{II} alcanza el final de la transcripción, el complejo THO carga el factor exportación Mex67 sobre el ARNm recién sintetizado (Strasser *et al.*, 2000a; Gwizdek *et al.*, 2006) y el ARNm, rodeado de multitud de factores de exportación (RNPM). El complejo THO transfiere el factor de exportación Mex67 al ARNm y es degradado por el exosoma nuclear, que reconoce el marcaje de ubiquitinación de Hpr1 (Strasser *et al.*, 2001; Kim *et al.*, 2004; Gwizdek *et al.*, 2006). Con ayuda de el factor Mtr2, Mex67 facilita el transporte específico de el RNPM a través del canal central del complejo de poros de la membrana nuclear (NPC) (Hurt *et al.*, 2000) (Figura 4).

En resumen, el complejo THO de *S. cerevisiae* protagoniza la conexión funcional entre los sucesivos pasos de la expresión génica a través de la biogénesis de

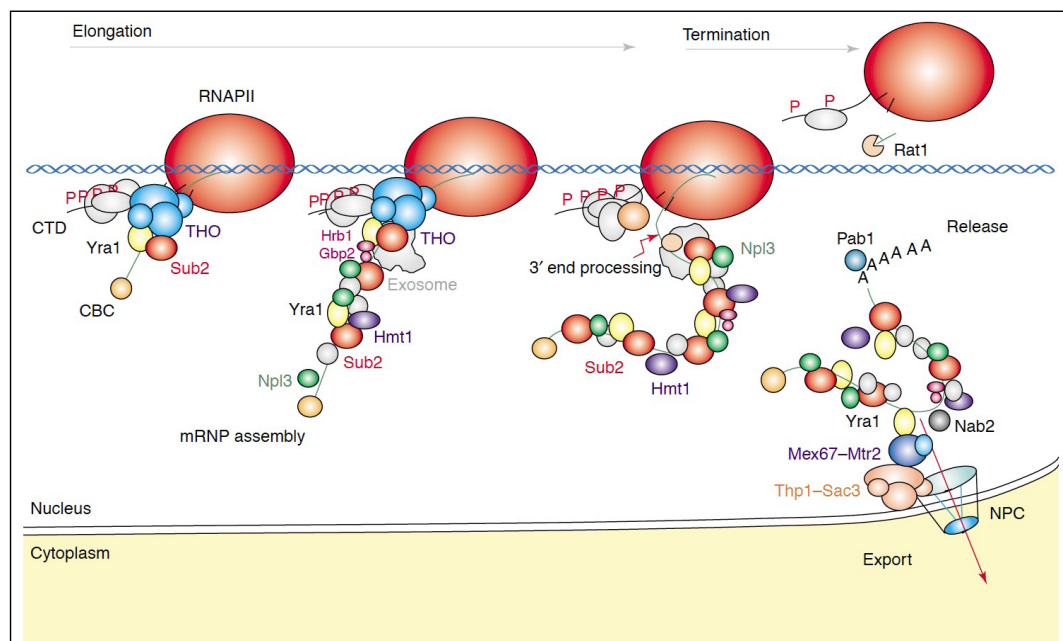


Figura 4: **Modelo de acción del complejo THO en *Saccharomyces cerevisiae*.** El esquema representa la evolución de la elongación transcripcional en *S. cerevisiae*. El complejo THO (azul claro) interacciona directamente con la ARNP_{II} (color rojo) y recorre el gen junto a ella (doble hélice de ADN representada en azul oscuro) facilitando la interacción de factores de exportación con el ARNm recién sintetizado (biogénesis de RNPM). En la terminación transcripcional, la ARNP_{II} se separa del ADN y el ARNm es liberado y dirigido hacia el NPC (azul) para alcanzar finalmente el citoplasma (amarillo). (Aguilera, 2005).

RNPm al tiempo que agiliza la expresión génica, facilita la exportación del ARNm y previene la inestabilidad genética asociada a la transcripción. Además, colabora con los mecanismos de control de calidad que garantizan la producción de moléculas de ARNm fiables y contribuye a la supervivencia celular en condiciones desfavorables (Fan *et al.*, 2001; Wellinger *et al.*, 2006; Luna *et al.*, 2008; Gómez-González *et al.*, 2009a; Jimeno *et al.*, 2010).

1.8. Microscopia electrónica y procesamiento digital de imágenes

La microscopía electrónica (ME) es una de las principales técnicas para el estudio estructural de macromoléculas biológicas. Frente a la cristalografía de rayos X y la resonancia magnética nuclear (RMN), la ME permite tratar especímenes grandes y flexibles, como proteínas multiméricas o grandes complejos macromoleculares. Este tipo de estructuras resultan inabordables por RMN, ya que esta técnica impone restricciones de tamaño (Pervushin *et al.*, 1997) y en muchas ocasiones también quedan fuera del alcance de la cristalografía, debido a la dificultad que presentan a la hora de formar cristales (Frank, 2006). La ME permite en ocasiones resolver estructuras a un nivel de resolución atómico, cuando éstas se ordenan en cristales bidimensionales o filamentos helicoidales en los que el elevado grado de periodicidad permite filtrar el ruido y recuperar la información de la estructura tridimensional (Henderson *et al.*, 1990; Valpuesta *et al.*, 1990). Sin embargo, en la mayoría de los casos, las proteínas o los complejos no interaccionan entre sí para dar lugar a estructuras cristalinas, sino que se distribuyen en la preparación como partículas individuales. En estos casos el análisis estructural permite alcanzar reconstrucciones tridimensionales de baja o media resolución (entre 30 y 10 Å), aunque se obtienen mejores resultados con partículas que presentan una elevada simetría, como por ejemplo los virus icosaédricos (Bottcher *et al.*, 1997b; Luque *et al.*, 2009). Todas las reconstrucciones tridimensionales presentadas en esta tesis doctoral han sido obtenidas por el método de análisis de partículas individuales sin imposición de simetría durante el proceso de reconstrucción (Jiang *et al.*, 2005; Llorca, 2005; Wang *et al.*, 2006a; Frank, 2009).

Habitualmente las imágenes tomadas a través del microscopio electrónico de transmisión presentan una relación señal ruido (SNR, signal to noise ratio) muy baja. Por ello es necesario combinar la señal proveniente de multitud de partículas idénticas para conseguir incrementar esta relación y alcanzar así mayor resolución en las reconstrucciones (Llorca, 2005). Sin embargo, en la práctica se observa que las partículas procedentes de una misma muestra no son idénticas y a menudo presentan variabilidad posicional y conformacional sobre la superficie de la rejilla. Así, en el análisis de partículas individuales por ME en primer lugar se seleccionan las partículas correspondientes al espécimen que se quiere analizar y a continuación se clasifican bidimensionalmente para contemplar la variabilidad presente en las partículas registradas. Una vez terminada la clasificación bidimensional y con una idea preliminar acerca del complejo que queremos reconstruir, comienza la reconstrucción tridimensional (Llorca, 2005; Frank, 2006) (Figura 5).

Para entender el proceso de reconstrucción tridimensional por ME es importante tener claro que todas las imágenes registradas en el microscopio electrónico de transmisión corresponden a proyecciones bidimensionales de muestras tridimensionales. Es decir, los electrones emitidos por el microscopio electrónico atraviesan los especímenes que se hallan en la superficie de la rejilla y las micrografías (placas fotográficas con emulsión fotosensible) registran sus proyecciones. Así, cada partícula es una proyección del espécimen en una orientación determinada. En las micrografías la información tridimensional del espécimen queda registrada en forma de niveles de gris (densidad óptica). Cuando las micrografías se digitalizan la densidad óptica se traduce a valores numéricos, de forma que la imagen queda representada por una matriz de datos en la que cada píxel tiene un valor correspondiente al nivel de gris registrado en la micrografía original. De esta forma, el procesamiento de imagen consiste en el análisis computacional de estos valores numéricos correspondientes a cada partícula.

Para poder obtener toda la información referente a la estructura tridimensional del espécimen es importante contar con el mayor rango posible de ángulos de proyección del espécimen. Por ello, en ocasiones se utiliza una inclinación de la rejilla en el microscopio electrónico (goniómetro) que favorece la variabilidad de

las proyecciones generadas. Durante el proceso de reconstrucción tridimensional se clasifican las partículas correspondientes a los diferentes ángulos de proyección y se comparan con las proyecciones generadas por un volumen de referencia en un proceso conocido como refinamiento angular iterativo (apartado 3.3.3 de Materiales y Métodos). El resultado de la comparación entre las proyecciones registradas en las micrografías (partículas) y las obtenidas a partir de un volumen de referencia es la asignación de ángulos a las proyecciones originales. Finalmente, teniendo los ángulos correspondientes a cada proyección, se retro-proyectan las partículas para obtener un volumen tridimensional correspondiente a la muestra del espécimen en cuestión (Frank, 2006) (Figura 5).

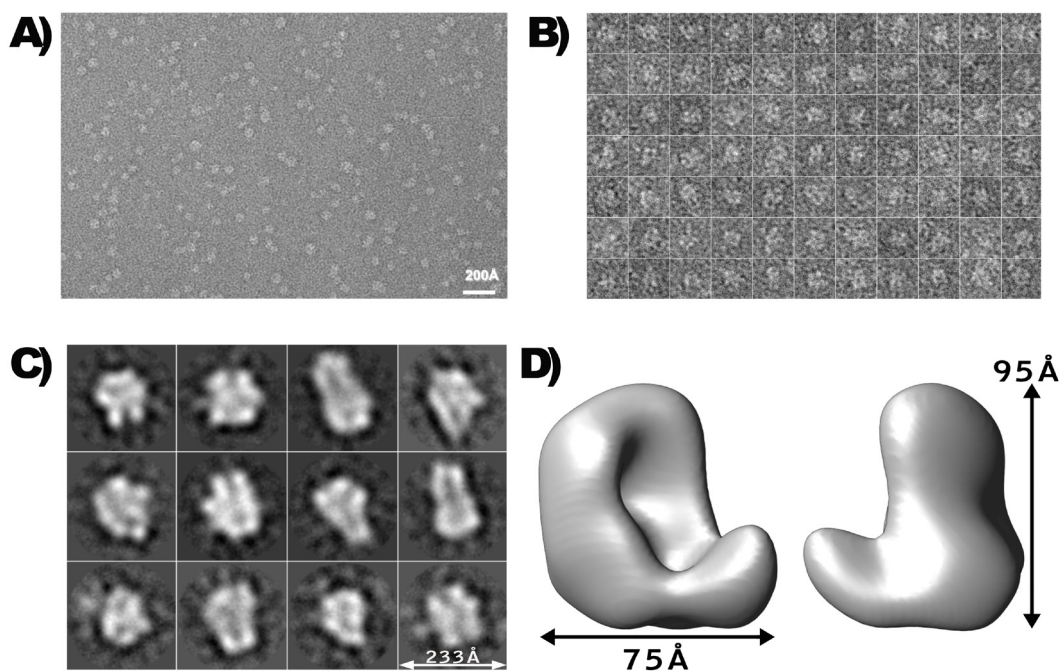


Figura 5: Etapas principales del procesamiento de partículas individuales por ME. **A)** Micrografía electrónica. Para analizar especímenes por el método de análisis de partículas individuales, en primer lugar se observa y la muestra a través del microscopio electrónico de transmisión y se registran sus proyecciones en micrografías. **B)** Partículas individuales. Las distintas proyecciones (partículas individuales) correspondientes al espécimen se seleccionan y se extraen de las micrografías generando un conjunto de imágenes cortadas en cajas cuadradas de la misma dimensión. Cada imagen (partícula) corresponde a una proyección de la misma macromolécula (especimen). **C)** Clasificación bidimensional. El conjunto de imágenes se analiza mediante clasificación bidimensional para contemplar la variabilidad de proyecciones y hacerse una idea del tamaño y la forma general del complejo. **D)** Reconstrucción tridimensional. Finalmente se reconstruye tridimensionalmente el espécimen y se genera un mapa de densidad en 3D. Los datos presentados en esta figura corresponden al análisis estructural del complejo Hsp40:RepE por microscopía electrónica de transmisión. Las partículas mostradas fueron cortadas en cajas de 100 x 100 píxeles a un muestreo de 2,33 Å/px. Imágenes cedidas por el doctor J. Cuéllar del Centro Nacional de Biotecnología.

1.9. Importancia del estudio estructural del complejo THO

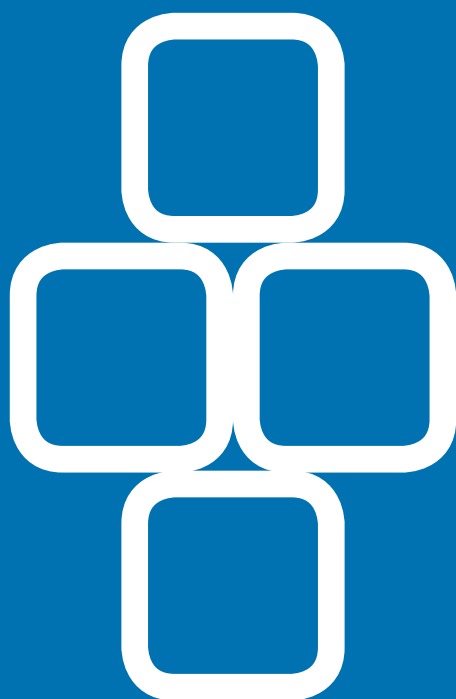
La expresión génica conlleva una serie de procesos que empiezan en la activación de genes y termina en la síntesis de proteínas. En células eucariotas este proceso tiene lugar en dos partes, la primera en el núcleo y la segunda en el citoplasma celular. Hasta la fecha los procesos que dirigen la síntesis citoplasmática de proteínas se han descrito con bastante precisión e incluso, en el año 2009, la caracterización estructural de los ribosomas y su función en las últimas fases de la expresión génica llevó a los doctores V. Ramakrishnan, T.A. Steitz y A.E. Yonath a ganar el Premio Nobel de Química (Williamson, 2009). Por el contrario, los procesos que ocurren en la primera fase de la expresión génica, dentro del núcleo, han sido menos caracterizados y algunos de los mecanismos que dan lugar a la producción y exportación núcleo-citoplasmática de información genética continúan siendo desconocidos para la comunidad científica actual (Ritchie *et al.*, 2009; Jimeno *et al.*, 2010).

En los últimos años se ha visto que algunos de los procesos implicados en las primeras fases de la expresión génica son de gran importancia para la supervivencia celular en determinadas condiciones. La función del complejo THO es esencial para la respuesta al estrés celular en *D. melanogaster*, para la supervivencia de embriones de ratón (Yang *et al.*, 2008) y para el desarrollo normal de la médula ósea y la hematopoyesis de mamíferos (Mancini *et al.*, 2010). El complejo THO de metazoos se ha visto además implicado en el desarrollo temprano del cáncer de mama (Guo *et al.*, 2005), de vejiga (Yu *et al.*, 2007) y de células pequeñas de pulmón (Yang *et al.*, 2008) y ha sido propuesto como una posible diana para tratamientos anti-cancerígenos (Li *et al.*, 2007). El complejo THO está también relacionado con la diarrea secretora aguda, una enfermedad bacteriana común que provoca una alta mortalidad de animales jóvenes y humanos (Niewold *et al.*, 2010).

El primer paso hacia la solución de enfermedades consiste en la comprensión de los procesos celulares que las causan. El análisis estructural de complejos macromoleculares contribuye frecuentemente a la comprensión de la función fisiológica que desempeñan las biomoléculas en su medio natural y aporta información decisiva acerca de las interacciones entre proteínas. Este tipo de aproximación con-

tribuye así a la comprensión de las redes de interacción proteica celulares y facilita una visión global de los procesos biológicos fundamentales. Sorprendentemente, ninguno de los componentes del complejo THO ha sido caracterizado estructuralmente hasta la fecha. De todos los complejos THO que han sido descritos sólo se conoce la estructura de un pequeño fragmento (13% del total) de la proteína humana Thoc1 (Niraula, 2005) homóloga de Hpr1. Desafortunadamente, este fragmento corresponde al dominio carboxi-terminal que parece ser el menos influyente en la función de la proteína y no se encuentra conservado a lo largo de la evolución, lo que hace suponer que no es representativo de la función general que desempeña esta proteína (García-Rubio *et al.*, 2008). Así, se puede considerar que no existe ningún dato estructural relevante acerca del complejo THO en ninguno de los organismo en los que ha sido descrito.

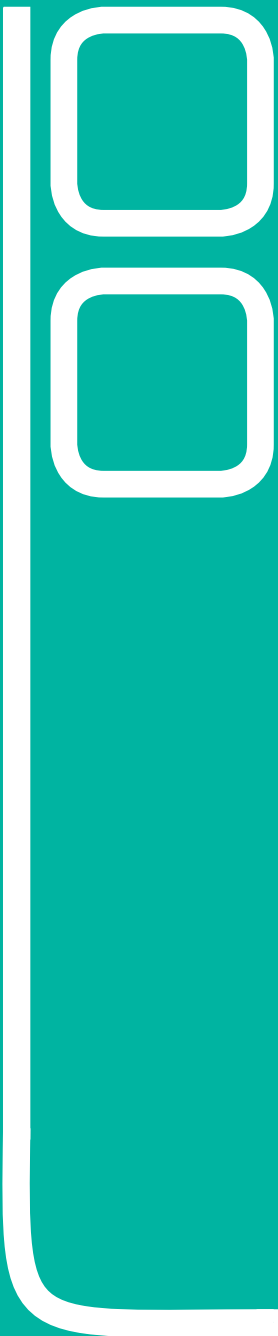
Esta tesis doctoral se centra en el estudio estructural del complejo THO de *S. cerevisiae* y muestra, entre otras cosas, la primera estructura tridimensional obtenida hasta el momento del complejo THO en su estado nativo. El trabajo pretende contribuir a la comprensión de las primeras fases de la expresión génica en organismos eucariotas. La caracterización estructural del complejo THO así como la descripción detallada de su modelo de acción y de las proteínas con las que interactúa pueden resultar determinantes a la hora de diseñar fármacos y tratamientos bioquímicos para potenciar la supervivencia celular en condiciones adversas.



2. Objetivos

Los objetivos planteados para la realización de esta tesis doctoral fueron los siguientes:

- 1.- Purificación del complejo THO de *Saccharomyces cerevisiae*.
- 2.- Reconstrucción tridimensional del complejo THO mediante microscopía electrónica de transmisión y procesamiento de imagen.
- 3.- Localización de los componentes del complejo THO y de alguna de las proteínas con las que interacciona para ejercer su función fisiológica *in vivo*.
- 4.- Caracterización de la posible interacción del complejo THO con los ácidos nucleicos ADN y ARN e identificación de las regiones involucradas en dicha interacción.



3. Materiales y Métodos

3.1. Obtención de muestras

3.1.1. Inserción de la cola TAP

La purificación en tándem por afinidad se realiza mediante el marcaje específico de una proteína diana con la secuencia TAP (*Tandem Affinity Purification*), de forma que se expresa la secuencia TAP a continuación de la proteína diana y se purifica ésta junto con las proteínas asociadas a ella en su forma nativa (TAP-tag) (Puig *et al.*, 2001). Así, el protocolo se basa en una secuencia genética diseñada específicamente para expresarse en las células junto a la proteína diana. La secuencia genética se añade a continuación de la secuencia del gen correspondiente a la proteína diana (Tsai *et al.*, 2006) y cuando se expresa la proteína diana, la secuencia TAP genera un péptido compuesto de tres segmentos: una zona de unión a calmodulina (CBP o *Calmodulin Binding Peptide*); un sitio de corte de la proteasa TEV (proteasa descrita inicialmente en el virus del tabaco, *Tobacco Etch Virus*) y una parte de la proteína A (proteína descrita inicialmente en *Staphylococcus aureus* que se utiliza en técnicas bioquímicas convencionales por su alta capacidad de interaccionar con inmunoglobulinas) (Figura 6).

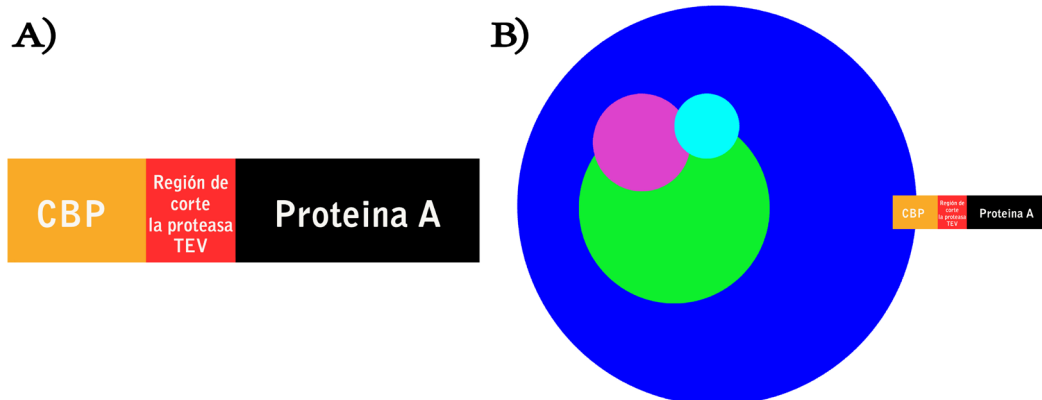


Figura 6: **Esquema del fragmento TAP y de su inserción en el complejo THO.** **A)** Cola TAP. El fragmento proteico TAP está compuesto por una región de interacción con calmodulina (CBP), un sitio de corte de la enzima TEV y una parte de la secuencia de la proteína A. **B)** Diseño de la cepa Tho2-TAP. En la cepa Tho2-TAP el fragmento TAP se insertó en el ADN de *S. cerevisiae* a continuación del gen que codifica para la proteína Tho2, la mas grande del complejo THO. De esta forma, tras la expresión y oligomerización del complejo el fragmento TAP se expresa a continuación del extremo carboxi-terminal de la proteína Tho2p, facilitando la purificación del complejo completo. Esta figura no mantiene las proporciones reales según el tamaño de cada componente.

Para purificar el complejo THO de *Saccharomyces cerevisiae* se modificaron genéticamente cuatro cepas añadiendo la secuencia TAP a cada una de ellas. En cada una de las cuatro cepas se añadió la secuencia TAP a continuación del extremo carboxi-terminal de cada uno de los cuatro componentes del complejo THO. Así, se obtuvieron cuatro cepas de *S. cerevisiae*, cada una diseñada para purificar el complejo THO utilizando uno de sus cuatro componentes como proteína diana. Las cuatro cepas se nombraron en función de la proteína marcada en cada caso: Tho2-TAP, Hpr1-TAP, Mft1-TAP y Thp2-TAP. A continuación se procedió a purificar el complejo a partir de cada una de las cepas. Para ello se crecieron las células en el medio de cultivo estándar YPD (10 gramos de Bacto-yeast extract, 20 gramos de BactoPeptone, 20 gramos de Dextrosa y agua miliQ hasta 1 litro) y se procedió a purificar el complejo siguiendo el protocolo de purificación en tándem por afinidad (Puig *et al.*, 2001; Burckstummer *et al.*, 2006; Tsai *et al.*, 2006). Las cepas Hpr1-TAP y Tho2-TAP permitieron purificar el complejo mientras que las cepas Mft1-TAP y Thp2-TAP apenas dieron resultados. Los mejores resultados se obtuvieron con la cepa Tho2-TAP, en la que la proteína mas grande del complejo (Tho2) había sido marcada en su extremo carboxi-terminal con la cola TAP (Figura 6). Estos experimentos fueron realizados por encargo a la compañía Euroscarf (*European Saccharomyces Cerevisiae Archive for Functional Analysis, Institute of Molecular Biosciences, Frankfurt*).

3.1.2. Cultivo de la cepa Tho2-TAP y purificación del complejo

Una vez realizado el marcaje con la secuencia TAP se procedió a la puesta a punto del protocolo de purificación a partir de la cepa Tho2-TAP. Atendiendo a la construcción de la cepa, la purificación se realizó utilizando la proteína Tho2 como diana. Todas las proteínas que se encuentren interaccionando de forma estable con Tho2 también copurificarán con la cola TAP (Puig *et al.*, 2001).

Al no existir ningún protocolo previo para la purificación del complejo THO, se comenzó siguiendo el protocolo estándar de purificación en tándem por afinidad y con la experiencia, se fueron adaptando las condiciones y los tampones de cada paso

de la purificación hasta optimizar el resultado de la purificación. A continuación se presenta el protocolo final utilizado para la purificación del complejo a partir de la cepa Tho2-TAP.

En primer lugar se preparó un precultivo a partir de la cepa Tho2-TAP en medio de cultivo YPD. Para ello se añadieron 0,5 ml de levaduras (almacenadas a 4 °C) a 200 ml de medio YPD y se incubaron durante aproximadamente 18h a 30 °C en agitación, hasta obtener una densidad óptica del cultivo próxima a 2. Para la incubación se utilizaron 16 botellas de 4 litros, cada una con 5 ml de precultivo y 2 litros de cultivo. La incubación se prolongó durante unas 20 horas a 30 °C hasta que los cultivos alcanzaron una densidad óptica superior a 2.

El volumen de cultivo se distribuyó en tubos y se centrifugó a 4500 revoluciones por minuto (rpm) durante 5 minutos a 4 °C para separar el medio de cultivo (en el sobrenadante) de las células (en el precipitado). Las células se disolvieron en agua mili-Q y se centrifugaron otra vez en las mismas condiciones para eliminar totalmente los restos del medio de cultivo. El precipitado de células se recogió, se pesó y se resuspendió en tampón A (10 mM Hepes pH 8,0, 250 mM NaCl, y 0,001% de β -mercaptoetanol) a razón de 20 ml de tampón por cada 100 gr de precipitado. Las células se congelaron por inmersión en nitrógeno líquido y se almacenaron a -80 °C. Partiendo de los 32 litros de cultivo se obtuvieron aproximadamente 450 gramos de células.

A continuación se rompieron las células con ayuda de hielo seco y una batidora convencional. La mezcla de levaduras y hielo seco se batió hasta obtener un polvo pastoso de células rotas. El polvo se diluyó en tampón B (40 mM Hepes pH 8,0, 250 mM NaCl, 1 mM DTT, 2 mM PMSF y dos cócteles de inhibidores de proteasas de Roche por cada 100 ml de tampón) a razón de 20 ml de tampón B por cada 100 gr de precipitado original.

La muestra se centrifugó dos veces, la primera a 20.000 rpm durante 20 minutos a 4 °C y la segunda durante una hora y media a 32000 rpm, descartando en ambos casos los precipitados y filtrando el sobrenadante a través de una gasa estéril. Los sobrenadantes se sometieron a tres horas de diálisis frente a tampón C (10 mM Hepes pH 8,0, 500 mM NaCl, 1 mM PMSF, 1 mM DTT y 20% de glicerol) a razón

de 1,5 litros de tampón C por cada 100gr de precipitado inicial. Así, la muestra se concentró, se homogenizó el tampón y se añadió glicerol. Las muestras fueron distribuidas en tubos Falcon de 50 ml y sumergidas en nitrógeno líquido para su congelación y almacenaje a -80°C .

Los tubos Falcon con las muestras se sumergieron en agua caliente para su descongelación y se les añadieron 1:1000 de Tritón X100 y 600ul de bolas de sefarosa de la compañía GE-Healthcare (*IgG Sepharose 6 Fast Flow beads*, número de catálogo 17-0969-01) a cada tubo Falcon, previamente lavadas y estabilizadas en tampón IPP 500 (0,5 M NaCl, 10 mM Tris-HCl pH 8,0 y 0,1% tritón X 100). Las muestras en los tubos Falcon con las bolas de sefarosa se incubaron durante toda la noche a 4°C en agitación suave constante.

Al día siguiente las muestras se centrifugaron durante 4 minutos a 3500 rpm y 4°C para precipitar las bolas de sefarosa y descartar el sobrenadante. Las bolas se resuspendieron de nuevo en tampón IPP 500, se dividieron en columnas de Bio-RAD (*Poly-prep chromatography columns*, número de catálogo 731-1550) y se lavaron primero con 30 ml de tampón IPP 500 y luego con otros 30ml de tampón TEV (0,5 M NaCl, 10 mM Tris-HCL pH 8,0, 0,5 mM EDTA, 1 mM DTT).

Una vez lavadas y estabilizadas las bolas de sefarosa en tampón TEV se añadió 1:1000 de enzima TEV a las columnas, previamente selladas para evitar fugas de líquido. La incubación de las bolas de sefarosa con la enzima TEV se dejó 2 horas a 16°C con agitación suave intermitente. Después de la incubación se recogió el excluido de las bolas de sefarosa en dos tandas de elución consecutivas. La muestra se diluyó dos veces en tampón IPP sin sal para disminuir la concentración salina a la mitad y se introdujo en una columna *Resource Q* de un mililitro para llevar a cabo la cromatografía de intercambio iónico mediante un sistema AKTA-purifier. Cuando la muestra terminó de pasar a través de la columna se pasó un gradiente de dos tampones similares, una con muy baja concentración de sal (tampón 1: 10mM Tris-HCl pH 8,0, 150mM NaCl) y otra con alta concentración de sal (tampón 2: 10mM Tris-HCl pH 8,0, 1M NaCl). La muestra de complejo THO eluyó de la columna a una concentración de $\pm 400\text{mM}$ de NaCl, dejando un tampón final con 10 mM de Tris-HCL pH 8,0 y $\pm 400\text{ mM}$ de NaCl. Las muestras obtenidas en la

cromatografía de intercambio iónico se cargaron en un gel desnaturalizante al poliacrilamida (SDS-PAGE) para comprobar su pureza, como se explica en el apartado 3.1.4.

3.1.3. Preparación de cepas mutantes

Para la realización de esta tesis doctoral se han utilizado cinco cepas de *S. cerevisiae* diferentes. La cepa Tho2-TAP, descrita en el apartado 3.1.1, se marcó con la secuencia TAP en el extremo carboxi-terminal de la proteína Tho2. Las otras cuatro cepas se explican a continuación:

- La cepa THO Δ Tex1 se hizo a partir de la cepa Tho2-TAP que se modificó genéticamente sustituyendo la secuencia correspondiente al gen TEX1 (codifica para la proteína Tex1) por la del gen de resistencia a higromicina. Al haberse diseñado esta cepa a partir de la cepa Tho2-TAP, también expresa la cola proteica TAP a continuación de la proteína Tho2, de forma que está preparada para la purificación en tándem por afinidad. Así, esta cepa sólo difiere de la cepa Tho2-TAP en que no expresa la proteína Tex1.

- La cepa Tho2-3C se modificó genéticamente introduciendo un sitio de corte de la proteasa 3C entre el aminoácido 1270 y el 1271 de la proteína Tho2. Esta cepa también se diseñó a partir de la cepa Tho2-TAP por lo que también presenta la cola TAP en el extremo carboxi terminal de la proteína Tho2. A partir de esta cepa se purificó el complejo THO y se incubó con la proteasa 3C para comprobar que el sitio de corte de la proteasa 3C quedaba accesible al medio y que el complejo THO se mantenía estable en ausencia del mismo (Figura 7).

- La cepa sh-Tho2 (shortened version of Tho2) se modificó genéticamente eliminando la región del gen THO2 correspondiente al extremo carboxi-terminal de la proteína Tho2, en base al sitio de corte observado en la cepa Tho2-3C. En esta cepa se sustituyó el extremo carboxi-terminal de la proteína Tho2, entre los aminoácidos 1270 y 1597 por el fragmento TAP, para permitir la purificación en tándem por afinidad del complejo THO truncado en la proteína Tho2. Así, esta cepa expresa la proteína Tho2 truncada en su extremo carboxi-terminal, que carece de los últimos 327 aminoácidos y tiene una longitud de 1270 aminoácidos en lugar de los 1597

originales. El fragmento TAP comienza inmediatamente después del aminoácido 1270 de la proteína, de forma que también está diseñada para la purificación en tándem por afinidad del complejo THO (Figura 7).



Figura 7: **Secuencia de la proteína Tho2 en las cepas Tho2-TAP, Tho2-C3 y sh-Tho2.** En la parte superior se muestra el esquema de la secuencia de la proteína Tho2 (azul) en la cepa Tho2-TAP, en la que la proteína tiene un total 1597 aminoácidos de longitud y el fragmento TAP después del aminoácido 1597 (extremo carboxi-terminal). En el centro se ha representado el esquema de la proteína Tho2 en la cepa Tho2-3C, que presenta un sitio de corte de la proteasa 3C (rojo) entre el aminoácido 1270 y el 1271. Finalmente se muestra el esquema de la proteína Tho2 en la cepa sh-Tho2, en la que los últimos 327 aminoácidos fueron sustituidos por el fragmento TAP, dejando la proteína Tho2 truncada en su extremo carboxi-terminal.

- La cepa $THO\Delta Tho2$ se modificó genéticamente sustituyendo la secuencia del gen $THO2$ por la del gen de resistencia a higromicina. La cepa $THO\Delta Tho2$ se diseñó en base a estudios previos que demuestran que en ausencia de la proteína Tho2 el complejo THO de *S. cerevisiae* no se ensambla y por tanto no es activo (Huertas *et al.*, 2006).

Estas construcciones fueron realizadas en el Instituto de Bioquímica y Biofísica de la Universidad de Varsovia en colaboración con los doctores Andrzej Dziembowski y Kamil Gewartowski.

3.1.4. Ensayos de electroforesis y *western-blot*

La electroforesis en condiciones desnaturalizantes se llevó a cabo en geles de poliacrilamida (SDS-PAGE) (Lange, 1970; Hayashi *et al.*, 1972). El gel concentrador se preparó a una concentración final de acrilamida del 3,5% en 285 mM Tris-HCl y pH 6,8. La concentración final de acrilamida en el gel separador fue del

10% en 370 mM Tris-HCl y pH 8,8. En ambos geles la concentración de SDS fue del 0,1%. Como iniciador de la polimerización se utilizó PSA (persulfato amónico) y como catalizador de la reacción de polimerización se utilizó TEMED (N,N,N', N' tetra-metil-etilen-diamina). La electroforesis se llevó a cabo en cubetas de la compañía GE Healthcare (*SE 250 Mini-Vertical Unit for two slab gels*, número de referencia 80-6147-45). Los geles tenían una anchura de 8 cm y una altura de 7 cm y un espesor de 0,75 mm. Antes de aplicar las muestras en el gel se les añadió tampón de carga 4X (62 mM Tris-HCl, pH 6,8, 8% glicerol, 2% SDS, 720 mM β -mercaptoetanol, 0,001% azul de bromofenol) y se desnaturalizaron calentándolas a 95 °C durante 5 minutos. La electroforesis se llevó a cabo en tampón de corrida (25 mM Tris-HCl, pH 8,3, 190 mM glicina y 0,1% SDS). Tras aplicar las muestras al gel, la electroforesis se inició a un voltaje constante de 80 V hasta que las muestras atravesaron el gel concentrador y a partir de ese momento el voltaje se aumentó a 120 V y se mantuvo constante hasta el final de la electroforesis. Se han utilizado marcadores de peso molecular de la compañía BioRad (*Precision Plus Protein Standards*, número de catálogo 161-0374). La tinción de geles realizada para la visualización de proteínas se hizo con tampón colorante azul de Coomassie (50% de Etanol, 01% Ácido Acético y 0.01% Coomassie azul brillante).

Para comprobar el resultado de las purificaciones en tándem por afinidad las muestras resultantes de la purificación se analizaron por *western-blot*. Para ello las muestras se cargaron en un gel SDS-PAGE y se transfirieron a una membrana de nitrocelulosa de 0,2 μ m (de la compañía BioRad, número de catálogo 162-0112) utilizando un sistema *Trans-blot SD Semi dry transfer cell* de la compañía BioRad (número de serie 221BR). Tanto la membrana como el gel se equilibraron con tampón de transferencia (25 mM Tris-HCl, pH 8,3, 124 mM glicina, 20% etanol y 10% SDS). La transferencia se realizó en módulos de papel de filtro de 7,5 x 10 cm de la compañía Bio-Rad (número de catálogo 1703965) a 200 mA y 10V durante una hora a temperatura ambiente. Una vez transferidas las proteínas, la membrana se incubó con solución de bloqueo (PBS 1X, 0,5% Tween 20 y 5% leche en polvo) durante una hora a temperatura ambiente en agitación suave. A continuación la membrana se incubó durante una hora con 1 ml de solución de bloqueo que contenía el anticu-

erpo específico contra el CBP de la cola TAP (anticuerpo anti-CBP producido por la empresa GenScript Corporation con el número de catálogo A00635). Después se lavó la membrana tres veces con solución de lavado (PBS 1X, 0,05% Tween 20 y 1% leche en polvo) y a continuación se incubó durante una hora con 5 ml de solución de bloqueo que contenía el anticuerpo secundario conjugado con peroxidasa (producido por la compañía Thermo Fisher Scientific, número de catálogo 32469). Por último se realizaron 3 lavados, cada uno con 100 ml de solución de lavado y una duración de 10 minutos y se reveló la membrana utilizando el sistema de quimio-luminiscencia ECL Plus (GE Healthcare, número de catálogo RPN2132) siguiendo las recomendaciones del fabricante. Para la identificación final de bandas luminiscentes se utilizaron películas Kodak “BioMax XAR film” de 13 x 18 cm (número de catálogo 165-1496). El anticuerpo anti-CBP mostró una banda clara a la altura de la proteína Tho2.

3.1.5. Diseño del anticuerpo y preparación inmunocomplejos

Para la preparación de inmunocomplejos se utilizó un anticuerpo policlonal de conejo encargado a la empresa Pacific-Immunology por el grupo de Andrés Aguilera (Universidad de Sevilla). El epítipo se eligió atendiendo al hecho de que el extremo carboxi-terminal de la proteína Hpr1 es susceptible de ubiquitinación durante el proceso de transcripción y que dicha región interacciona con el factor de exportación Mex67 *in vivo* (Gwizdek *et al.*, 2006; Hobeika *et al.*, 2007; Gómez-González *et al.*, 2009b). Así pues, atendiendo a las regiones propuestas para la ubiquitinación del extremo carboxi-terminal de Hpr1 se eligieron varios péptidos como posibles epítipos para anticuerpos y se encargaron los correspondientes anticuerpos para ser generados en conejo. El mejor resultado se obtuvo para el anticuerpo cuyo epítipo reconocía diecisiete aminoácidos localizados entre las posiciones 636 a 653 de la secuencia proteica de Hpr1 de *Saccharomyces cerevisiae* (la longitud total de la secuencia es de 752 aminoácidos). Se trata así de un anticuerpo policlonal generado en conejo contra un péptido de 17 aminoácidos (Figura 8).

MSNTEELIQNSIGFLQKTFKALPVSFDSIRHEPLPSSMLHASVLNFEWEPELEKNISAIHD
 RDSLDIILKRFIIDSMTNAIEDEEENLEKGLLNSCIGLDFVYNSRFNRSPASWGNTF
 FELFSTIIDLLNSPSTFLKFWPYAESRIEWFKMNTSVEPVS LGESNLISYKQPLYEKL RH
 WNDILAKLENNDILNTVKHYNMKYKLENFLSELLPINEESNFNRSSASISALQESDNEWNR
 SARERESNRSSDVIFAADYNFVFYHLIICPIEFAFS DLEYKNDVDRLSPLLDAILEIEE
 NFYSKIKMNNRTRYSL EEA LNT EY YANYDVMTPKLPVYMKHSNAMKMDRNEFWANLQNIK
 ESDDYTLRPTIMDISL SNTTCLYKQLTQEDDDYRKQFQLQCFTTNLIRNLISSDETRN
 FYKSCYLRENPLSDIDFENLDEVNKKRGLNLC SYICDNRVLFYKIKDPDFYRVIRKLMS
 SDEKFTTAKIDGKFEQNFRISEKIPPPAFDETFKKFTFIKMGKNLINNVWKIPTGLDK
 IEQEVKKPEGVYEEAAQAKWESKISSETSGGEAKDEIIRQWQTLRFLRSRYLDFDQVNEK
 TGV DGLFE EPRKVEALDDSFKEKLLYKINQEHKK **LQDAREYKIGKERKKRALEEEASF**
EREQKIKSQRINSASQTEGDELKSEQTQPKGEISEENTKISSEVSSQDPDPSGVAGEFAP
QNTTAQLENPKTEDNNAATSNISNGSSTQDMK

Figura 8: **Diseño del anticuerpo anti-Hpr1**. La imagen muestra la secuencia de aminoácidos completa de la proteína Hpr1. En color rojo se ha resaltado el epítipo de 17 aminoácidos que reconoce el anticuerpo anti-Hpr1.

La especificidad del anticuerpo anti-Hpr1 se comprobó mediante ensayos de *western-blot* y *dot-blot*. Para los ensayos de *western-blot* se utilizó un protocolo idéntico al descrito en el apartado 3.1.4 utilizando como anticuerpo primario el anti-Hpr1 en lugar del anti-CBP. Los ensayos de *western-blot* demostraron que el anticuerpo reconoce el epítipo de la proteína Hpr1 en su estado desnaturalizado, pero para el estudio estructural de inmunocomplejos por microscopía electrónica es necesario que el anticuerpo reconozca el epítipo de la proteína Hpr1 formando parte del complejo THO en su estado nativo. Para ello, se diseñó un experimento de *dot-blot* en el que las muestras de complejo THO recién purificado fueron depositadas en una membrana de nitrocelulosa (Bio-Rad número de catálogo 1703965) y antes de que se secasen del todo, se procedió a bloquear la membrana con solución de bloqueo (PBS 1X, 0,5% Tween 20 y 5% leche en polvo) durante una hora a temperatura ambiente en agitación suave. De esta forma se fijaron a la membrana de nitrocelulosa complejos THO que quedarían parcialmente desnaturalizados en lugar de aquellos procedentes de un gel SDS-PAGE que están totalmente desnaturalizados y con todos sus componentes separados por tamaño.

Una vez comprobada la eficacia, la especificidad y la capacidad de interacción del anticuerpo anti-Hpr1 con las muestras de complejo THO total y parcialmente desnaturalizadas, se procedió a la preparación de inmunocomplejos para el análisis por microscopía electrónica. Para ello los anticuerpos se incubaron con una

muestra recién purificada de complejo THO. En la incubación se utilizó una relación molar de 1:8 (THO:anticuerpo) manteniendo un exceso de anticuerpo en relación a la cantidad de complejos THO y aumentando así la probabilidad de interacción. La incubación se mantuvo durante veinte minutos a temperatura ambiente en agitación suave. Para el análisis estructural las muestras fueron diluidas hasta una concentración final de 6 ug/ml utilizando el tampón final de THO (10 mM de Tris-HCl pH 8,8 y 400 mM de NaCl) y aplicadas sobre la superficie de carbón de las rejillas de microscopia electrónica, como se explica en el apartado 3.6.

3.2. Cálculo estequiométrico

El estudio estequiométrico del complejo THO se realizó en colaboración con los doctores Kamil Gewartowski y Andrzej Dziembowski en el instituto de Bioquímica y Biofísica de la Universidad de Varsovia. Para ello se cargó una cantidad conocida de muestra en un gel desnaturalizante en gradiente de acrilamida comprado a la compañía Invitrogen (número de catálogo NP0323BOX). Una vez terminada la electroforesis el gel se tiñó con colorante azul de Coomassie y se escaneó en un escáner EPSON *perfection V750 PRO* atendiendo al histograma para registrar la mayor cantidad de tonalidades posibles en la imagen resultante. A continuación el gel se analizó con el programa de cálculo densitométrico Image Quant. Este procedimiento se realizó varias veces consecutivas utilizando diferentes geles y con muestras procedentes de distintas purificaciones con el objetivo de obtener una aproximación estadísticamente significativa.

3.3. Análisis de la interacción con ácidos nucleicos

Este trabajo fue realizado durante mi estancia en el instituto de Bioquímica y Biofísica de la Universidad de Varsovia, bajo la supervisión de los doctores Andrzej Dziembowski y Kamil Gewartowski.

Para comprobar la posible interacción entre el complejo THO de *S. cerevisiae* y los ácidos nucleicos se diseñó un experimento de entrecruzamiento con

ADN de cadena simple (ADNcs), ADN de cadena doble (ADNcd) y ARN. Para ello se mezclaron 50 μ l de complejo THO recién purificado con 150 μ l de tampón GET (10mM Tris-HCl pH 8,0, 15% glicerol y 1mM EDTA) y se añadieron los nucleótidos marcados radiactivamente. En los experimentos presentados tanto el ADN como el ARN tenían una longitud de 44 nucleótidos pero también se comprobó la interacción del complejo THO con cadenas nucleotídicas de mayor tamaño. La mezcla se incubó durante 5 minutos a temperatura ambiente. Después se colocó la muestra en el interior de un transiluminador de la compañía Hoefer (UVC 500 Ultraviolet Crosslinker, número de catálogo UVC500-230V) durante un minuto. Las muestras se diluyeron en tampón de carga de electroforesis y se sometieron a un choque térmico de cinco minutos a 95 °C para ser cargadas en un gel desnaturalizante al 10% de acrilamida proporcionado por la compañía Invitrogen (número de catálogo NP0343BOX). Las proteínas se visualizaron mediante tinción con tampón colorante azul de Coomassie (50% de Etanol, 0,1% Ácido Acético y 0,01% azul de Coomassie brillante) mientras que el ADN y el ARN marcado radiactivamente se detectaron utilizando un escáner de la compañía GE-Healthcare (Typhoon-FLA 7000, número de catálogo 28-9558-09).

3.4. Estudio fenotípico

El estudio fenotípico se realizó para las cepas Tho2-TAP (control positivo, produce el complejo THO activo), sh-Tho2 (produce el complejo THO con la proteína Tho2 truncada en su extremo carboxi-terminal), THO Δ Tho2 (control negativo, no produce complejo THO activo) y THO Δ Tex1 (produce complejo THO activo pero no la proteína Tex1) atendiendo a dos factores: el crecimiento celular y la capacidad de expresión del gen lac-Z. Los ensayos fueron realizados en colaboración con los doctores Kamil Gewartowski y Andrzej Dziembowski en el instituto de Bioquímica y Biofísica de la Universidad de Varsovia

Para estudiar el crecimiento celular se cultivaron las cuatro cepas a 30 °C y a 37 °C durante tres días. De cada cepa se hicieron cuatro diluciones con diferente cantidad de células, de forma que el cultivo mas concentrado presentaba una densidad óptica

de 1, el siguiente diez menor (0,1), el siguiente cien veces menor (0,01) y el último mil veces menor (0,001). Por otro lado, el estudio de la expresión del gen lac-Z se hizo atendiendo a la cantidad de β -galactosidasa (producto de expresión del gen lac-Z) producida a 30 °C y a 37 °C. Para ello se centrifugó un mililitro de cultivo de cada cepa y se le añadieron al precipitado 0,5 ml de tampón Z (100 mM Na₂HPO₄/NaH₂PO₄ pH 7,2, 10 mM KCl, 1 mM MgSO₄ y 350 μ l de β -mercaptoetanol) y 200 μ l de eter diluido en agua hasta saturación. La mezcla se agitó durante 20 segundos, se centrifugó de nuevo durante 1 minuto y se esperaron unos 10 minutos hasta que el eter se evaporó. La muestra se incubó 5 minutos a 30 °C ó a 37 °C, realizando el mismo experimento en paralelo para cada cepa a ambas temperaturas. A continuación se añadieron 100 μ l de ONPG (orto-nitrofenilgalactopiranosido) y se incubaron de nuevo a las mismas temperaturas hasta que mostraron un color amarillo. La reacción se paró añadiendo 250 μ l de 1M Na₂CO₃. Las muestras se centrifugaron de nuevo durante 5 minutos a 10.000 rpm y se cargaron 200 μ l de cada muestra en una placa de 96 pocillos para medir la densidad óptica a 420 nm con un lector de placas. La producción de β -galactosidasa se calculó con la siguiente fórmula:

$$\text{actividad} = (1000 \times \text{valor de densidad óptica}) / (\text{volumen} \times \text{valor de densidad óptica} \times \text{tiempo})$$

3.5. Predicción de estructura terciaria

La predicción computacional de estructura de la región amino-terminal de la proteína Tex1 de *S. cerevisiae* fue realizada en colaboración con los doctores Juan Carlos Sánchez y Florencio Pazos del Centro Nacional de Biotecnología. Para ello se empleó el método de reconocimiento de plegamiento (*Fold-Recognition* ó *Threading*) que se utiliza para modelar la estructura de proteínas que presentan un plegamiento similar a otras proteínas de estructura conocida y funciona utilizando el conocimiento estadístico de la relación entre las estructuras conocidas y la secuencia de las proteínas que se desea modelar (Bowie *et al.*, 1991; Jones *et al.*, 1992; Jones *et al.*, 1999). Para el cálculo de la estructura se utilizaron los siguientes programas: BLAST (Altschul *et al.*, 1990), MUSCLE (Edgar, 2004b; a), JALVIEW (Waterhouse *et al.*, 2009), PHYRE (Kelley *et al.*, 2009), I-TASSER (Zhang, 2008; Roy *et*

al., 2010), Swiss-PDBviewer (Guex *et al.*, 1997; Guex *et al.*, 2009), PYMOL (DeLano, 2002) y UCSF-CHIMERA (Pettersen *et al.*, 2004; Goddard *et al.*, 2007).

3.6. Microscopía electrónica

3.6.1. Preparación de muestras de tinción negativa

Una vez purificados los complejos y comprobada la pureza de los mismos por análisis en SDS-PAGE, las muestras se prepararon para ser observadas a través del microscopio electrónico de transmisión mediante tinción negativa. Para ello, en primer lugar se depositó manualmente una película de carbón sobre rejillas de Cobre/Rodio de 300 agujeros (Maxtaform HR25, Cu/Rh EM grids) por flotación. Unos segundos antes de añadir la muestra a la rejilla, ésta se sometió a una descarga iónica durante treinta segundos, lo que provoca que la película de carbón se haga más hidrofílica y favorezca la fijación de la muestra a la superficie de la rejilla. Terminada la descarga, la rejilla se incubó con la muestra durante aproximadamente tres minutos a temperatura ambiente.

El volumen de muestra utilizado para la incubación de cada rejilla es de aproximadamente 5 μ l y la concentración ha varía entre 1 y 2 μ M, según las diferentes muestras. Tras la incubación la rejilla se secó ligeramente utilizando papel de filtro fino de la compañía Whatman (número 1, referencia de catálogo 1001-090) dejando únicamente una fina película de muestra acuosa sobre la rejilla. Después la rejilla se incubó con acetato de uranilo al 2% (peso/volumen) durante un minuto. A continuación se retiró el exceso de agente de tinción utilizando un papel de filtro idéntico al anterior y la rejilla se dejó secar al aire libre a temperatura ambiente durante 30 minutos antes de su visualización en el microscopio electrónico de transmisión (Figura 9).

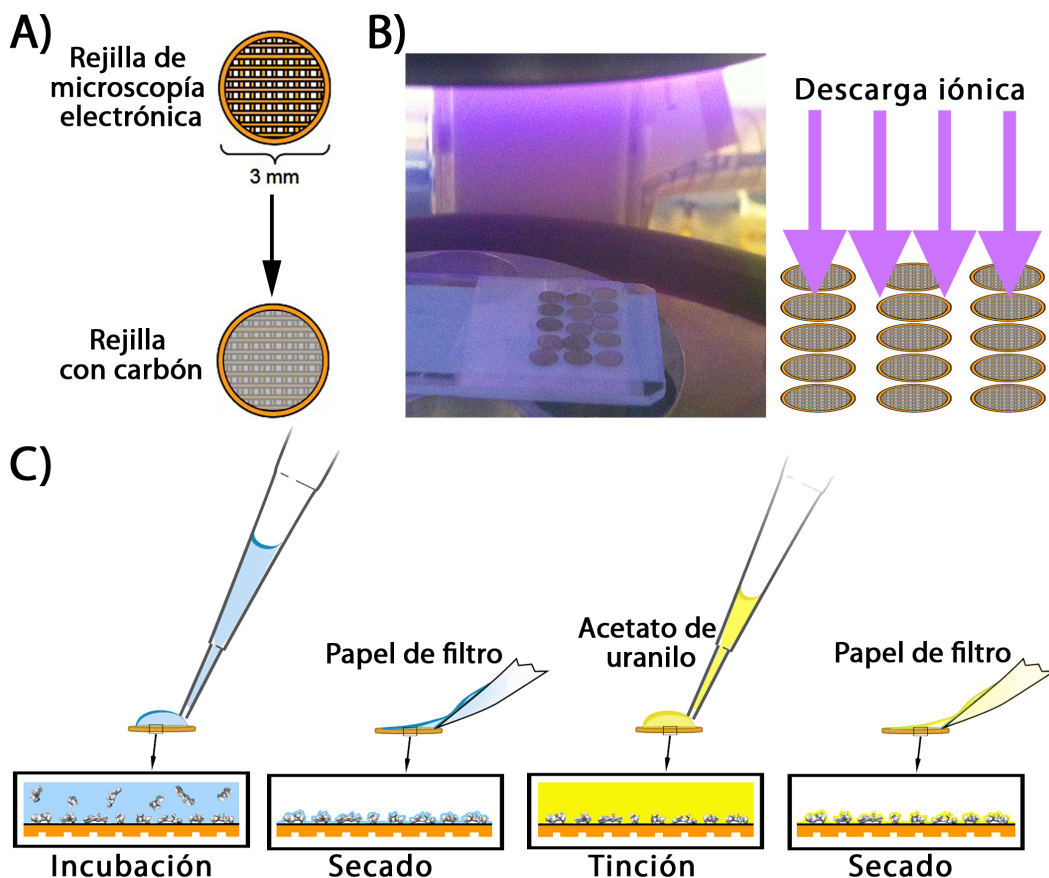


Figura 9: Preparación de rejillas para microscopía electrónica. El esquema muestra los pasos realizados para la preparación de muestras de microscopía electrónica por tinción negativa en rejillas cubiertas con carbón. A) En primer lugar las rejillas de microscopía electrónica son cubiertas con una fina capa de carbón que se distribuye en paralelo a la superficie principal de la rejilla. B) A continuación las rejillas cubiertas son sometidas a una descarga iónica que provoca que la superficie de carbón se haga mas hidrofílica, facilitando así la adhesión de las moléculas biológicas a la superficie de la rejilla. C) A continuación se añade cuidadosamente la muestra directamente sobre la rejilla utilizando una pipeta de alta precisión y se incuba durante ± 3 minutos. La muestra se seca parcialmente con ayuda de un papel de filtro y se incuba de nuevo (± 1 minuto) con el agente de tinción (acetato de uranilo al 2%). Finalmente se elimina el exceso de agente de tinción de la superficie de la rejilla con ayuda de un papel de filtro y se deja secar la rejilla (± 30 minutos). Debajo de cada paso se ha representado una ampliación de la superficie de la rejilla en la que se distinguen los complejos en cada paso de la preparación.

3.6.2. Adquisición de imágenes

Las rejillas de tinción negativa de los diferentes especímenes se observaron en un microscopio electrónico de transmisión JEOL 1200EX-II equipado con un filamento de wolframio previamente estabilizado a 100 KV. Las micrografías se tomaron a 60.000 aumentos a 0 y a 30° de inclinación del goniómetro, lo que permitió aumentar la diversidad de imágenes de proyección presentes en las micrografías. Todas las micrografías electrónicas se tomaron en condiciones de mínima dosis (aproximadamente 10 e-/Å² sobre la muestra) utilizando película ortocromática Kodak-electron SO-163 de 6,5 cm de ancho x 9 cm de largo (número de catálogo 167-9273). Todos los procesamientos presentados en este trabajo han sido obtenidos a partir de una sola rejilla de cada muestra y procurando utilizar la mínima cantidad de zonas de cada rejilla, para disminuir el la variabilidad existente entre distintas rejillas y diferentes zonas de la misma rejilla.

3.7. Procesamiento de partículas individuales

3.7.1. Digitalización y cálculo de la Transformada de Fourier

Para digitalizar las micrografías se utilizó un microdensitómetro Photoscan TD Zeiss-Intergraph utilizando un tamaño de paso de 7 micras, lo que determina un tamaño de píxel de 2,33 Å/píxel para todos los casos. Antes de llevar a cabo la selección de partículas se calculó la transformada de Fourier para cada micrografía y se utilizó esta información como criterio para descartar las micrografías astigmáticas.

3.7.2. Selección de partículas, corrección de CTF y análisis en 2D

La selección de partículas se hizo de manualmente utilizando el paquete informático XMIPP (Marabini *et al.*, 1996). Para el procesamiento del complejo THO:Tex1 se seleccionaron un total de 13.273 partículas, para el de THOΔTex1 14.115 partículas y para el de THO:antiHpr1 un total de 8.250 partículas.

Una vez seleccionadas las partículas de cada procesamiento se procedió a la corrección de la CTF (*Contrast Transfer Function*). Para ello se calcularon los parámetros de la CTF utilizando el programa CtfFind incluido en XMIPP y después se realizó la corrección de la CTF invirtiendo el signo de aquellas fases que mostraban signo negativo, utilizando el comando *bctf* del paquete informático BSOFT (*Bernard's Software Package*) (Heymann, 2001; 2007) o de forma automatizada con la última versión del paquete informático XMIPP (Scheres *et al.*, 2008; Sorzano *et al.*, 2010). A continuación se extrajeron las partículas seleccionadas de las micrografías en cajas de 140x140 píxeles para las reconstrucciones de THOΔTex1 y THO:Tex1 y de 180x180 para la reconstrucción del complejo THO:Tex1:antiHpr1, utilizando en todos los casos la última versión del paquete de informático XMIPP.

Para la clasificación en dos dimensiones (2D) las partículas fueron reescaladas hasta un tamaño final de 70x70 píxeles, utilizando el comando *xmipp_scale* del paquete informático XMIPP. La clasificación bidimensional se hizo de manera independiente utilizando los procedimientos de máxima verosimilitud, redes neuronales auto-organizativas y análisis de componentes principales cuyas características se explican a continuación.

El procedimiento de máxima verosimilitud realiza el alineamiento bidimensional multireferencia con el objetivo de asignar las partículas a un número determinado de clases. Este recurso está disponible en el paquete informático XMIPP a través del programa *ml_align2d* (Maximum-likelihood alignment in 2D) (Scheres *et al.*, 2005a; 2005b). Este algoritmo de clasificación maximiza la probabilidad de haber observado los datos experimentales dado un modelo estadístico que incluye además de las imágenes promediadas, una descripción formal del ruido. La maximización se lleva a cabo mediante el algoritmo de *expectation-maximization* en el que las decisiones de alineamiento y asignación no se toman en función de pequeñas diferencias en el valor de correlación cruzada, sino que se calcula la probabilidad para todas las posibles asignaciones (Scheres *et al.*, 2007). Cada imagen media que se genera tiene en cuenta la probabilidad de asignación de cada partícula a esa clase. Para generar el conjunto inicial de imágenes medias que sirven como referencias las partículas iniciales no alineadas se dividen aleatoriamente en grupos y se calcula la

imagen media para cada grupo. Para optimizar el funcionamiento de este método de clasificación todas las partículas fueron normalizadas previamente con el programa *normalize* de XMIPP.

La clasificación bidimensional basada en redes neuronales auto-organizativas se realizó mediante el programa *KerDenSOM* (*Kernel Density Estimator Self-Organizing Map*) implementado en el paquete informático XMIPP (Scheres *et al.*, 2008). Este procedimiento está basado en el algoritmo SOM (Self-Organizing Maps), que establece una correspondencia entre la información de entrada y un espacio de salida bidimensional, identificando relaciones desconocidas previamente entre los miembros del conjunto de datos (Kangas *et al.*, 1990; Kohonen, 1990). El algoritmo *KerDenSOM* presenta la ventaja de que la función de densidad de probabilidad se estima de forma no paramétrica, de manera que los vectores representantes tienden a poseer la misma distribución estadística que los datos originales (Pascual-Montano *et al.*, 2001). Así, este algoritmo alinea las partículas y luego las separa según las diferencias encontradas en un área determinada establecida previamente por el usuario (máscara) (Frank *et al.*, 1992; Penczek *et al.*, 1992). Este algoritmo a sido particularmente útil a la hora de separar las partículas del complejo THO con anticuerpo (anti-Hpr1) de aquellas sin anticuerpo.

El procedimiento de análisis de componentes principales permite obtener imágenes promedio libres de referencia y es el método de clasificación bidimensional implementado en el paquete informático EMAN (Ludtke *et al.*, 1999). El análisis de componentes principales es una técnica estadística de síntesis de la información o reducción de la dimensión (número de variables). Es decir, en un banco de datos con muchas variables (las partículas), se reducirá el número de variables perdiendo la menor cantidad de información posible. El objetivo de este análisis estadístico es identificar el número de factores subyacentes explicativos que comparte el conjunto de datos y que representen la variabilidad observada entre ellos, que en el caso de las imágenes de microscopía electrónica se hace atendiendo a la correlación entre las partículas (Daffertshofer *et al.*, 2004). Así, el programa divide el conjunto de imágenes y las somete a un proceso de refinamiento multi-referencia que consiste en calcular los armónicos radiales y rotacionales para cada una de las partículas,

para después clasificarlas en función de sus armónicos. De esta forma se obtienen en primer lugar entre diez y veinte clases que representan la variabilidad existente en el conjunto de imágenes, de manera simplificada. A continuación se alinean y promedian las diferentes partículas dentro de cada clase y este esquema se repite iterativamente para finalmente alinear las imágenes medias entre sí y llevar a cabo un proceso equivalente al análisis estadístico multi-variante que se denomina SVD (*Singular Value Decomposition*). Este proceso permite identificar las variables más significativas que representan la variabilidad del conjunto de imágenes medias. A continuación se aplica una nueva clasificación más precisa en la que las partículas se separan en un número final de clases determinado por el usuario. Por último las partículas dentro de cada clase se promedian y se genera un nuevo conjunto de imágenes medias. El proceso se repite normalmente entre 8 y 10 veces, produciendo finalmente un conjunto de imágenes medias que representan las diferentes clases que existen en el conjunto inicial de partículas.

3.7.3. Reconstrucción tridimensional

3.7.3.1. Obtención de volúmenes de referencia

El proceso de reconstrucción tridimensional se realiza a través de lo que se conoce como refinamiento angular iterativo. Este proceso comienza con la proyección de un volumen de referencia (modelo inicial) que da como resultado un conjunto de imágenes bidimensionales llamadas proyecciones teóricas. Cuando se pretende reconstruir un espécimen del que existen datos estructurales previos a veces es posible utilizar estos datos como volumen de referencia. Por el contrario, cuando se analiza un espécimen del que no existe ningún dato estructural previo, se realiza la reconstrucción tridimensional partiendo de un volumen de referencia generado computacionalmente. Para la realización de esta tesis doctoral se han empleado diferentes métodos para obtención de volúmenes de referencia. Con el objetivo de comprobar que el resultado final obtenido en cada refinamiento es independiente del volumen de referencia utilizado se han generado diferentes volúmenes y se han re-

alizado varios refinamientos en paralelo hasta comprobar que el resultado final era el mismo en todos los casos. Los volúmenes de referencia se han generado utilizando el paquete informático EMAN, atendiendo principalmente a tres metodologías diferentes:

- El método de líneas comunes permite generar volúmenes de referencia a partir de una clasificación bidimensional previa de los partículas. Así, se seleccionan varias imágenes medias con una relación señal-ruido alta y se analizan en el espacio de Fourier. El método se basa en el hecho de que las transformadas de Fourier de las proyecciones de un mismo volumen tridimensional constituyen planos del espacio de Fourier tridimensional del volumen al que pertenecen. De esta forma, a través del programa StartAny implementado en el paquete informático EMAN, se puede combinar la información de las diferentes medias bidimensionales de partículas en el espacio de Fourier y buscar así las líneas comunes que comparten en un espacio tridimensional. Una vez encontradas las líneas comunes se interpola el resto de la información para obtener un volumen a baja resolución que se filtra y se puede utilizar como volumen de referencia para el refinamiento angular iterativo.

- Otra forma de obtener volúmenes de referencia consiste en generar modelos iniciales partiendo de dimensiones predefinidas computacionalmente. El programa `makeinitialmodel.py` implementado en el paquete informático EMAN permite generar volúmenes tridimensionales a partir de dimensiones establecidas por el usuario. El resultado es un volumen con una forma geométrica que recibe el nombre de *blob*. Así, un *blob* es un volumen tridimensional que no guarda ninguna semejanza con el espécimen que se pretende reconstruir y que se utiliza exclusivamente para generar un conjunto de proyecciones con las que comparar las partículas para comenzar un refinamiento angular iterativo.

- Por último, se puede generar volúmenes de referencia a partir de ruido. Este método consiste en utilizar como volumen de referencia un conjunto de ruido distribuido en las tres dimensiones del espacio. La opción *noise* del programa *proc3d* implementado en el paquete informático EMAN permite aplicar ruido en tres dimensiones a cualquier volumen previamente existente. Así, se generan volúmenes de tipo *blob* y después se les ha añadido ruido de tal forma que la señal del ruido

sea tan intensa como la señal de volumen utilizado. El resultado es un volumen tridimensional compuesto exclusivamente por ruido distribuido aleatoriamente en el espacio tridimensional que puede ser utilizado como volumen de referencia para iniciar el proceso de refinamiento angular iterativo.

3.7.3.2. Refinamiento angular iterativo

La reconstrucción tridimensional de estructuras por microscopía electrónica se hace en base a ciclos de proyección, comparación, asignación de ángulos y retroproyección, en lo que se conoce como refinamiento angular iterativo. El proceso comienza con la proyección de un volumen de referencia o modelo inicial (apartado anterior) que genera las llamadas proyecciones teóricas. Las proyecciones teóricas son imágenes bidimensionales que provienen de la proyección computacional del volumen de referencia y se caracterizan por presentar una alta relación señal ruido y un valor correspondiente al ángulo de proyección conocido. El número de proyecciones teóricas dependerá del valor angular utilizado para la proyección del volumen de referencia, de forma que cuanto menor sea el ángulo de proyección, mayor será el número de proyecciones generadas. A continuación las proyecciones teóricas se comparan con el conjunto de partículas seleccionadas y extraídas de las micrografías, a las que llamamos proyecciones experimentales. Las proyecciones experimentales son imágenes bidimensionales que provienen de la proyección del espécimen biológico y que han sido registradas en las micrografías a través del microscopio electrónico. Las proyecciones experimentales contienen información tridimensional del espécimen biológico real y se caracterizan por presentar una baja relación señal ruido ya que provienen directamente de la observación de muestras a través microscopio electrónico (Figura 10).

Así se realiza una clasificación bidimensional en la que cada proyección experimental (partículas) es asignada a una proyección teórica, en función de su coeficiente de correlación. La comparación de las proyecciones teóricas con las proyecciones experimentales da como resultado la asignación de ángulos de proyección a las proyecciones experimentales en función del valor angular de la proyec-

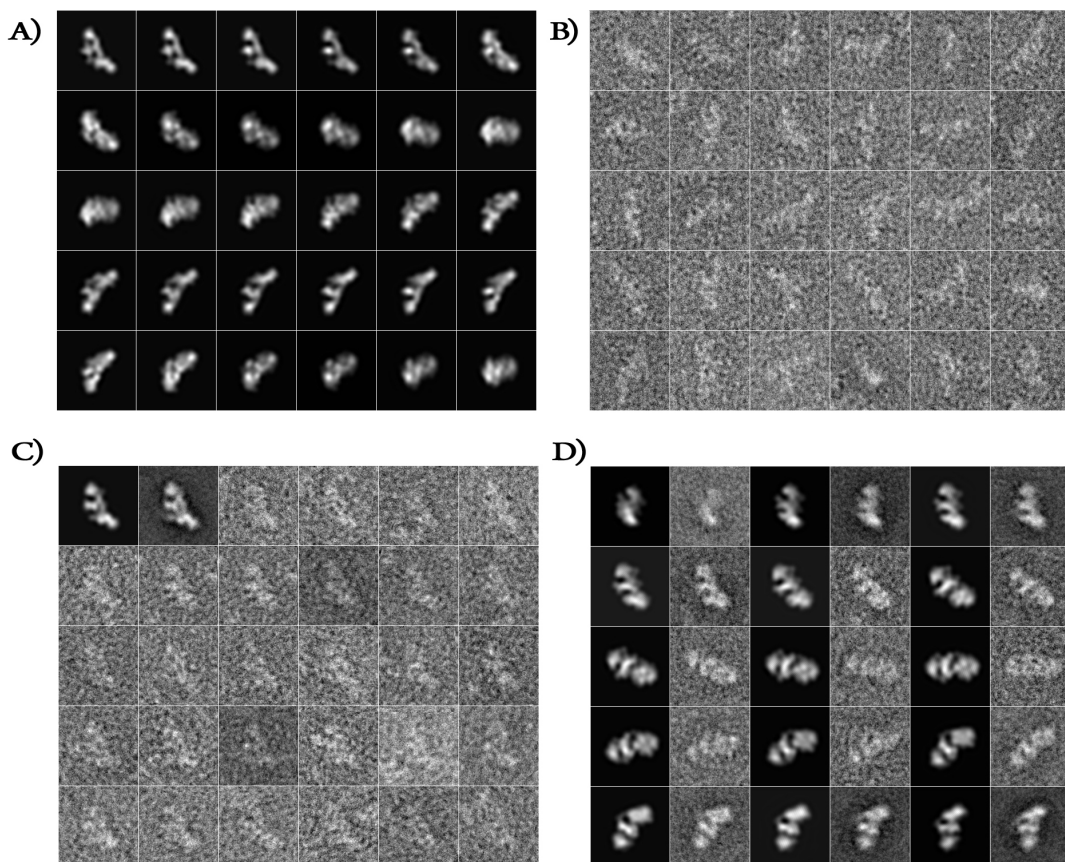


Figura 10: Proyecciones teóricas y proyecciones experimentales. **A)** Proyecciones teóricas obtenidas a partir de la “proyección” computacional del volumen de referencia utilizado para un refinamiento angular iterativo. **B)** Proyecciones experimentales (partículas) registradas en micrografías a través de un microscopio electrónico de transmisión y digitalizadas a 2,33 Å/px. Las proyecciones experimentales corresponden a proyecciones de la muestra biológica real. **C)** En orden desde arriba a la izquierda hacia la derecha y abajo: proyección teórica, imagen promedia de partículas asignadas y visión parcial de partículas asignadas a la proyección. La figura muestra las partículas a las que se ha asignado el ángulo correspondiente a la proyección teórica (primera imagen). Las partículas asignadas a cada proyección se promedian para aumentar la relación ruido-sígnal (segunda imagen). **D)** Proyecciones teóricas e imágenes promedio de las proyecciones experimentales asignadas. A pesar de que la relación ruido señal aumenta al promediar las partículas, la relación ruido señal de la proyección teórica siempre es mucho mayor, puesto que proceden de un modelo tridimensional inicial proyectado computacionalmente.

ción teórica a la que han sido asignadas. Así, todas las proyecciones experimentales asignadas a una misma proyección teórica comparten el mismo valor de ángulo de proyección y se promedian para obtener una imagen con mayor señal ruido. Con la clasificación bidimensional se obtiene una media de partículas asignadas a cada proyección teórica del volumen inicial. Finalmente se procede a la retroproyección de las medias obtenidas para cada proyección teórica, utilizando para ello el valor del ángulo de proyección asignado a ésta. De esta forma se obtiene un nuevo modelo tridimensional procedente de las partículas (proyecciones experimentales). El

modelo resultante será mejor cuanto mejor sea el volumen de referencia utilizado, dado que la correlación entre las proyecciones teóricas del volumen de referencia y las partículas será mas alta y la asignación angular será mas precisa. El modelo obtenido en cada ciclo de refinamiento angular iterativo se filtra y se utiliza como volumen de referencia para el siguiente ciclo de refinamiento.

Cada ciclo de proyección, comparación, asignación angular y retroproyección constituye una iteración y para la reconstrucción de volúmenes hacen falta generalmente varios cientos de iteraciones, por lo que es necesaria una capacidad de cálculo computacional muy elevada. Para la obtención de las reconstrucciones presentadas en esta tesis doctoral se han utilizado varios superordenadores adaptados con nodos de memoria compartida (*cluster*). A continuación se cita el nombre y la localización de los superordenadores que han sido mas relevantes para el desarrollo de este trabajo, en orden alfabético: CRUNCHY (Centro Nacional de Biotecnología, Madrid); FINIS-TERRE (Centro de Supercomputación de Galicia, Santiago de Compostela); JUMILLA (Centro Nacional de Biotecnología, Madrid); MARENOSTRUM (Centro Nacional de Supercomputación, Barcelona); TRUENO (Secretaría General Adjunta de Informática del CSIC, Madrid).

El refinamiento angular iterativo y la reconstrucción tridimensional de volúmenes ha sido realizada mediante los paquetes informáticos EMAN, XMIPP y SPIDER. A continuación se explican las características principales de cada uno de ellos:

El paquete informático EMAN (Ludtke *et al.*, 1999) realiza el refinamiento angular iterativo de forma semi-automatizada a través del programa refine. Este programa resultó especialmente útil en las primeras fases de la reconstrucción tridimensional de todas las estructuras presentadas, debido a que minimiza el sesgo introducido por el volumen de referencia. Esto se consigue a través del sistema de independencia de volumen de referencia o IMB (Initial Model Bias) que se consigue a través del alineamiento y promediado múltiple de las imágenes asignadas a cada proyección teórica. Es decir, el programa refine genera las proyecciones del volumen de referencia (proyecciones teóricas), las compara con las partículas (proyecciones experimentales) y obtiene una primera clasificación en la que las partículas se

alinean contra la proyección teórica correspondiente. A continuación las partículas asignadas a cada proyección teórica se someten a ocho ciclos de alineamiento y promediado, de forma que el primer ciclo se realiza a través de un alineamiento de las partículas con la proyección teórica y los demás se realizan alineando las partículas contra su propia media, generada en el ciclo anterior. Así se consigue que la imagen final que se utilizará para la retroproyección sea mas independiente del volumen de referencia, puesto que procede de un alineamiento contra su propia media y no contra la proyección del volumen de referencia (proyección teórica).

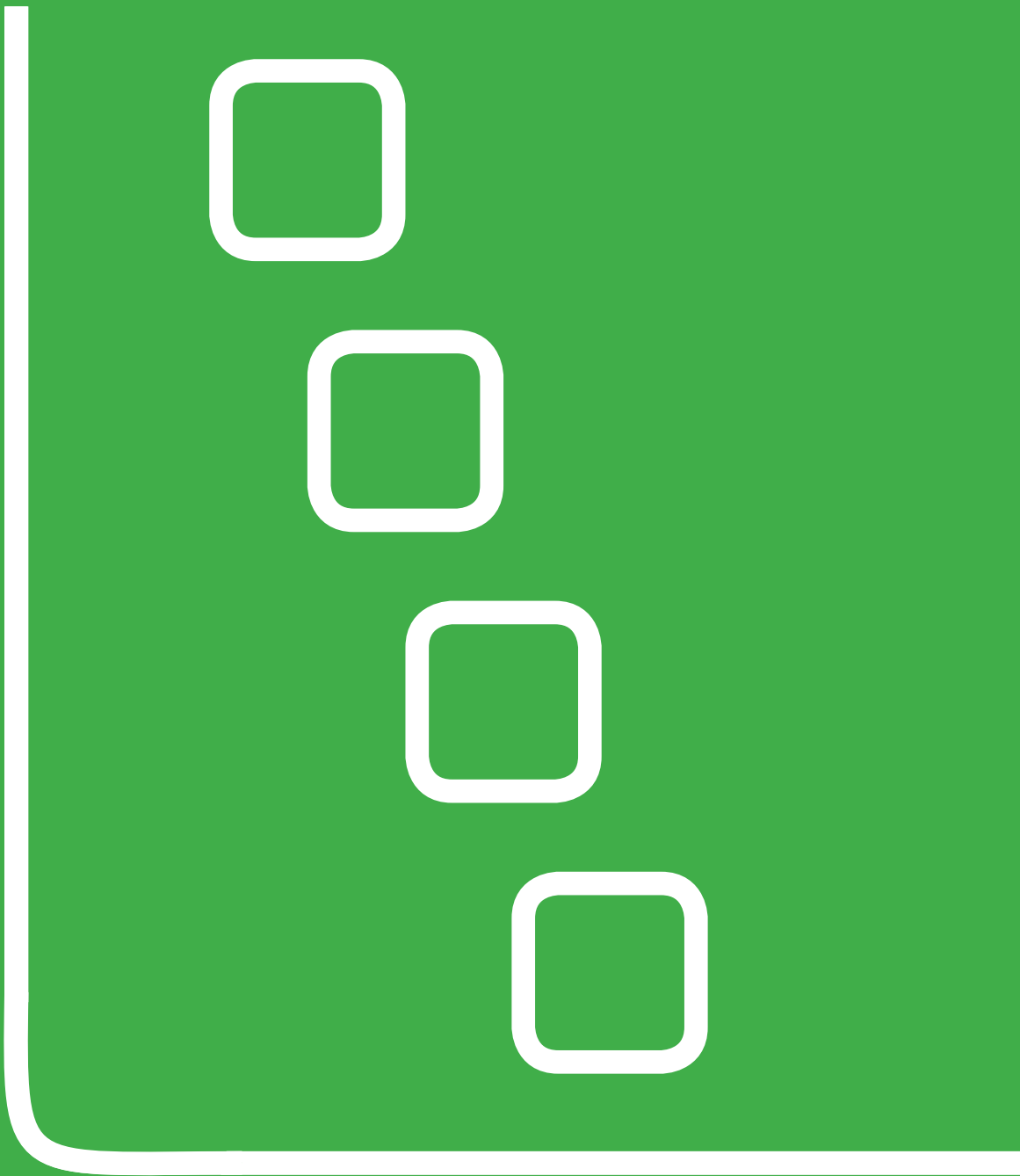
La última versión del paquete informático XMIPP (Scheres *et al.*, 2008; Sorzano *et al.*, 2010) ha sido de gran utilidad en las fases mas avanzadas de la reconstrucción tridimensional. El comando *xmipp_protocols* ejecuta un gestor de procesamiento que permite avanzar en el procesamiento de forma semi-automática e intuitiva, desde la selección de partículas en las micrografías hasta la reconstrucción tridimensional. El programa *Projection Matching* realiza el refinamiento angular iterativo comenzando con un algoritmo que busca la mejor orientación angular para cada imagen. Para ello cada partícula (proyección experimental) es comparada con todas las proyecciones generadas por el volumen de referencia (proyecciones teóricas) a través de un algoritmo específico que busca el coeficiente de correlación mas alto. Así, la asignación de ángulos a las partículas depende directamente del coeficiente de correlación entre estas y las proyecciones teóricas. Finalmente se realiza la retroproyección y se obtiene un nuevo modelo, que se utilizará como volumen de referencia de la siguiente iteración. Este procedimiento presenta mayor dependencia de volumen de referencia que el implementado en el paquete informático EMAN, por lo que funciona mejor cuando se utiliza con volúmenes de referencia buenos, que mantienen una estabilidad y fiabilidad con respecto al espécimen original. El algoritmo del programa *Projection Matching* permite adquirir mayor resolución en los volúmenes finales y genera resultados mejores que los obtenidos con EMAN.

El paquete informático SPIDER (Frank *et al.*, 1996; Shaikh *et al.*, 2008) también ha sido de gran utilidad en las fases mas avanzadas de la reconstrucción tridimensional. A diferencia de EMAN y XMIPP, el refinamiento angular iterativo realizado con SPIDER se ha ejecutado paso a paso de forma totalmente manual.

Este tipo de abordaje resulta mucho mas laborioso que los anteriores pero permite un control mas preciso de cada paso que se realiza en el refinamiento. La versión utilizada ha sido la 10.0 para la que estaban disponibles algunos de los *scripts* que permiten personalizar cada paso del refinamiento (entre ellos, *params.spi*, *good001.spi*, *refproj.bat*, *apmq.bat*, *angles.bat*, *align.bat*, *b16.cla*, *b17.avg*, *b20.tot*, *b26.bpr*, *b29.stk*, *b30.trs*, *b96.sin* y *generasel.bat*) alcanzándose de esta forma los mejores resultados finales.

3.7.3.3. Cálculo de resolución

Una vez concluido el proceso de refinamiento angular iterativo se lleva a cabo la estimación de la resolución para la estructura resuelta mediante el paquete informático SPIDER o EMAN. El objetivo de esta estimación consiste en definir de manera cuantitativa la resolución que se aprecia en el volumen obtenido. Para ello el conjunto inicial de partículas se divide en dos grupos y se calcula la reconstrucción tridimensional de cada grupo. Después las dos estructuras se comparan en el espacio de Fourier y se calcula la correlación entre ambas en función de la frecuencia espacial o FSC (*Fourier Shell Correlation*) (Saxton *et al.*, 1982). Existen diferentes criterios para definir qué valor de correlación es el que mejor refleja la resolución del volumen obtenido. Todos las resoluciones de las reconstrucciones presentadas en esta tesis doctoral se han calculado atendiendo al valor de la frecuencia espacial correspondiente a un valor de la FSC de 0,5 (Bottcher *et al.*, 1997a) y de 0,3 (Orlova *et al.*, 1997). Los volúmenes finales presentados se han filtrado a la resolución estimada para eliminar las frecuencias espaciales accesorias.



4. Resultados

4.1. Purificación del complejo a partir de la cepa Tho2-TAP

Para realizar esta tesis doctoral se purificó el complejo THO a partir de un cultivo celular de *S. cerevisiae*. Para ello se introdujo el fragmento TAP en el extremo carboxi-terminal de cada una de las cuatro proteínas del complejo y se procedió a purificar el complejo a partir de cada una de las cepas. El mejor resultado se obtuvo en la purificación del complejo a partir de la cepa Tho2-TAP, en la que la proteína Tho2 expresaba el fragmento TAP a continuación de su extremo carboxi-terminal (apartado 3.1.1 de Materiales y Métodos). Una vez obtenida esta cepa el complejo se purificó siguiendo estrictamente el protocolo de purificación en tándem por afinidad (Puig *et al.*, 2001) pero el rendimiento obtenido era muy bajo.

Se intentó sobreexpresar el complejo *in vivo* en las células para lograr un mayor rendimiento de la purificación pero tras varios intentos los resultados de la purificación no mejoraban. Se optó por adaptar el protocolo de purificación para procesar una cantidad mayor de cultivo inicial de levaduras y con la experiencia se fueron ajustando las condiciones y los tampones para optimizar los resultados de la purificación. Así, en el protocolo final de purificación se partía de 32 litros de cultivo, se eliminó el paso de purificación con bolas de calmodulina y se añadió un paso de cromatografía de afinidad (apartado 3.1.2). De esta forma se consiguió optimizar la purificación hasta conseguir aproximadamente 75 µg de complejo puro a partir de 500 gramos de células (32 litros de cultivo).

Una vez finalizada la purificación, se observó que la muestra era muy inestable y tendía a degradarse después de tres o cuatro días a partir de su purificación. Se probaron varios tipos de inhibidores de proteasas pero el resultado resultó ser el mismo. Se probó a utilizar tubos de baja adherencia para evitar que el complejo interaccionase con el plástico de los tubos, pero ningún intento mejoró la estabilidad del complejo. Se probó también a distribuir las muestras en tubos muy finos con el objetivo de disminuir la superficie de contacto entre la muestra y el aire, pensando que quizá fuese éste el causante de la degradación, pero el resultado continuaba inalterable. Se optó por cambiar el tipo de tampón y las condiciones de éste, se añadió glicerol y se modificó la cantidad de sal y de otros componentes pero el complejo

seguía degradándose en los tres o cuatro días siguientes a su purificación.

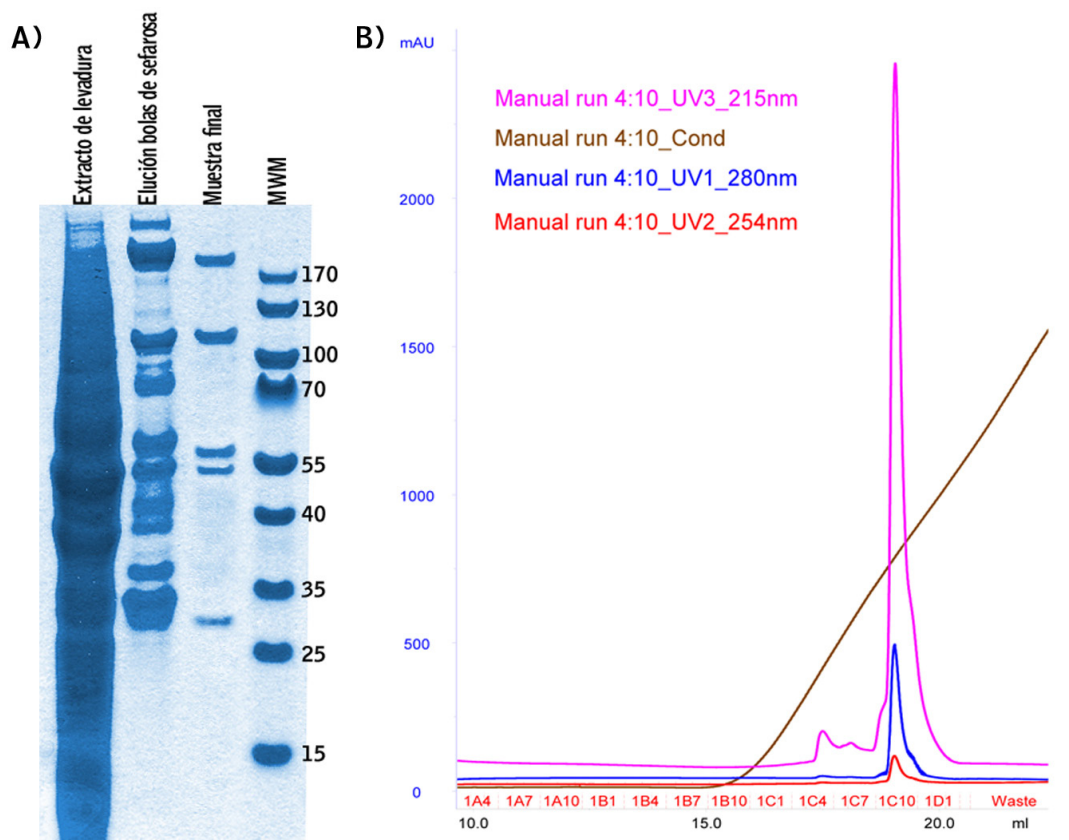
La muestra final se obtenía a una concentración de entre 15 y 30 microgramos por mililitro, lo cual dificultaba su utilización para algunas técnicas bioquímicas. Se intentó concentrar la muestra utilizando el sistema de concentración por presión de la compañía Amicón pero se perdía gran parte de la muestra, como si el complejo quedase pegado a los filtros. Se probó a utilizar el mismo sistema sin presión y con diferentes filtros concentradores, pero tampoco dio resultado. La muestra se intentó también concentrar mediante centrifugación en tubos diseñados específicamente para ello y también por diálisis pero el complejo sufría daños rápidamente y tendía a degradarse por lo que no fue posible alcanzar concentraciones mayores de 30 – 35 $\mu\text{g}/\mu\text{l}$.

Se intentó almacenar el complejo purificado a -20 y -80 °C pero se observó que el complejo no resistía los procesos de congelación y descongelación. Se probaron varios métodos de congelación, como la inmersión en nitrógeno líquido, en nieve carbónica o colocando la muestra directamente en congeladores, pero tras la descongelación el complejo aparecía totalmente degradado y la muestra descongelada era inutilizable. Se intentó proteger la muestra utilizando diferentes tampones y cantidades crecientes de glicerol y de sacarosa, llegando a concentraciones cercanas al 80%, pero tras la descongelación el complejo aparecía siempre degradado.

Las dificultades encontradas en la inestabilidad del complejo han sido la mayor complicación metodológica encontrada durante la realización de esta tesis doctoral. Después de todas las condiciones probadas se concluyó que el tampón que mejor funcionaba era el que menos cantidad de solutos tenía, por lo que todos los análisis estructurales y bioquímicos presentados en esta tesis doctoral se han realizado en un tampón final que contenía exclusivamente 10 mM de tampón Tris-HCl a pH 8,0 y 400 mM NaCl. Todas las muestras se purificaron siguiendo el protocolo descrito en el apartado 3.1.2 y se almacenaron en hielo a 4 °C durante el tiempo que permanecían estables.

El resultado de la purificación del complejo a partir de la cepa Tho2-TAP se cargó en un gel SDS-PAGE al 10% de acrilamida (apartado 3.1.4). Sorprendente-

mente, se observó que la muestra contenía cinco proteínas que se reflejaban como cinco bandas en el gel (Figura 11).



RESULTADOS

Figura 11: **Resultados de la purificación a partir de la cepa Tho2-TAP.** **A)** La imagen muestra cuatro carriles de un gel SDS-PAGE al 10% de acrilamida. En el primer carril (izquierda) se cargaron 10 μ l del extracto de levadura a partir de cual se comenzó la purificación. A continuación se cargaron 20 μ l de la muestra resultante de la elución de las bolas de sefarosa (protocolo de purificación en el apartado 3.1.2). En el tercer carril se cargaron 40 μ l de las fracciones 1C8 a 1C11 resultantes de la cromatografía de intercambio iónico (el cromatograma correspondiente aparece en la figura B). En el último carril del gel se cargaron 7 μ l de marcadores de peso molecular (MwM). A la derecha de los marcadores de peso molecular se muestran los tamaños correspondientes a cada banda. En la figura se observa que en la purificación en tándem por afinidad con la cola TAP expresada en el extremo carboxi-terminal de la proteína Tho2 es posible aislar cinco proteínas entre las cuales la mayor supera los 170 kDa de tamaño y la menor ronda los 30 kDa. **B)** La imagen muestra el resultado de la cromatografía de intercambio iónico en una columna *Resource Q* realizada como paso final de la purificación en tándem por afinidad a partir de la cepa Tho2-TAP. En el gráfico se presenta la absorción a 215, 280 y 254 nanómetros de longitud de onda (colores rosa, azul y rojo respectivamente) y la conductividad (color marrón), que aumenta a lo largo de la cromatografía con el incremento de fuerza iónica. El pico del cromatograma comprendido entre las fracciones 1C8 y 1C11 corresponde al tercer carril del gel presentado en la figura A. La concentración salina del tampón en el momento de la elución del complejo es de aproximadamente 400 mM.

4.2. Identificación de la proteína Tex1 asociada al complejo THO

El complejo THO de *S. cerevisiae* está compuesto por cuatro proteínas (apartado 1.6). La observación de cinco bandas en el gel llevó al análisis de cada una de ellas por espectroscopía de masas MALDI-TOF. Para ello las bandas fueron cortadas con una cuchilla estéril, separadas del gel y almacenadas en tubos eppendorf de 1,5 ml para ser entregadas al servicio de Proteómica del Centro Nacional de Biotecnología. Además de la confirmación de la presencia de las cuatro bandas correspondientes a las cuatro proteínas del complejo THO (Tho2, Hpr1, Mft1 y Thp2) el resultado del análisis mostró que la quinta banda del gel correspondía a la proteína Tex1 de *S. cerevisiae* (Figura 12).

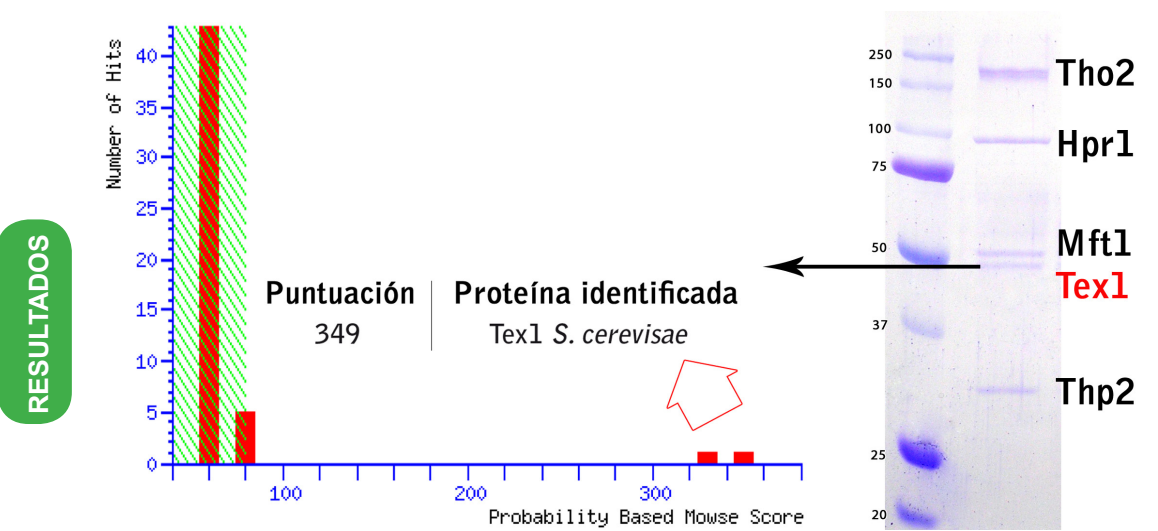


Figura 12: **Identificación de Tex1 de *S. cerevisiae***. La figura muestra los resultados del análisis por espectroscopía de masas MALDI-TOF de la banda cuarta banda (empezando por arriba) del gel que se muestra a continuación. Como se ve en la figura, la identificación es inequívoca puesto que sólo existen dos valores de puntuación altos, ambos correspondientes a la proteína Tex1 de *S. cerevisiae*. El resto de proteínas identificadas que aparecen a la izquierda de la imagen, dentro del rectángulo rallado en verde, representan los valores no significativos de la identificación, con puntuaciones bajas. El gel que aparece a la derecha muestra el resultado de la purificación del complejo a partir de la muestra Tho2-TAP en el que se han representado las proteínas correspondientes a cada banda, identificadas todas por espectroscopía de masas MALDI-TOF. A la izquierda del gel se han representado los tamaños correspondientes a cada banda de los marcadores de peso molecular (MwM) expresados en kDa.

Este resultado mostró que cuando se utiliza una cola TAP en el extremo carboxi-terminal de la subunidad Tho2 del complejo THO, la proteína Tex1 copurifica junto a los cuatro componentes del complejo THO, lo que sugiere una interacción estable entre Tex1 y el complejo THO de *S. cerevisiae* en condiciones nativas.

La proteína Tex1 de *S. cerevisiae* tiene un tamaño de 47 kDa (422 aminoácidos) y fue descrita en el año 2002 como un factor de transcripción y exportación de ARNm que interacciona de forma estable y copurifica con el complejo THO (Jimeno *et al.*, 2002; Strasser *et al.*, 2002). Tex1 es una proteína de función desconocida, muy conservada a lo largo de la evolución y se ha comprobado que su homóloga en humanos, Thoc3, también interacciona de forma estable con el complejo THO *in vivo* (Masuda *et al.*, 2005; Cheng *et al.*, 2006). De hecho, la proteína Tex1 interacciona con el complejo THO de *D. melanogaster*, *H. sapiens* y *S. cerevisiae* incluso en condiciones de alta salinidad y se ha propuesto como uno de los factores mas íntimamente asociado con el complejo THO. La interacción de Tex1 con el complejo THO está muy conservada a lo largo de la evolución y parece contribuir a la realización de su función fisiológica (Strasser *et al.*, 2002; Rehwinkel *et al.*, 2004; Saguez *et al.*, 2005; Jimeno *et al.*, 2010). Así pues, el resultado obtenido en la purificación del complejo THO es consistente con los datos bibliográficos previos.

4.3. Reconstrucción tridimensional del complejo THO:Tex1

Una vez obtenido el complejo THO:Tex1 como resultado de la purificación a partir de la cepa Tho2-TAP se procedió a su caracterización estructural por microscopía electrónica y procesamiento de imagen. Para ello se prepararon rejillas de tinción negativa (apartado 3.6.1) a partir de las fracciones de mayor pureza (Figura 11) de la muestra recién purificada. Se prepararon diluciones comprendidas entre 1:4 y 1:6 a partir de la muestra original utilizando el mismo tampón en que se encontraba el complejo tras el último paso de purificación. Las rejillas fueron observadas en un microscopio electrónico de transmisión JEOL 1200EX-II (apartado 3.6.2). A primera vista se diferenciaron a través de los binoculares del microscopio electrónico unas formas alargadas y cóncavas dispersas por el campo de visión. Las partículas aparecían distribuidas homogéneamente y presentaban unos patrones constantes de forma y tamaño con una tinción uniforme. Los complejos parecían depositarse sobre la superficie de la rejilla en una posición mayoritaria, con el eje mas largo de la partícula en el plano paralelo a la superficie de la rejilla (Figura 13).

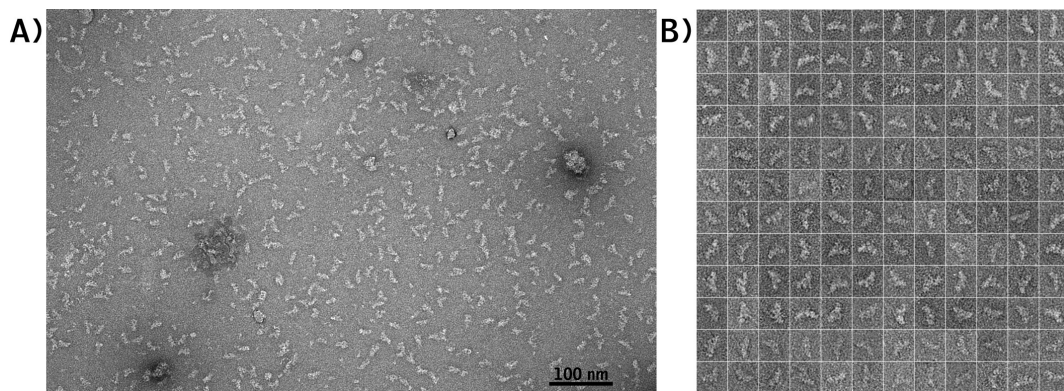


Figura 13: **Observación y selección de partículas del complejo THO:Tex1.** **A)** Micrografía. La imagen muestra un campo de tinción negativa de una micrografía correspondiente a la muestra obtenida a partir de la cepa Tho2-TAP. La imagen fue tomada a 60.000 aumentos y en ella se pueden apreciar una serie de estructuras mas o menos uniformes que se repiten de forma constante en todo el campo de visión: el complejo THO:Tex1. Abajo a la derecha se muestra la barra de escala. **B)** Selección de partículas individuales. Vista parcial del conjunto de partículas seleccionadas manualmente y cortadas en imágenes individuales de 140 x 140 píxeles. Cada partícula corresponde a una proyección del complejo THO:Tex1 en una posición determinada. Se observa que el conjunto de imágenes es bastante uniforme. Se aprecia una vista mayoritaria en la que el complejo muestra una forma alargada y curvada, que recuerda a la forma de un *croissant*.

Las partículas del complejo se seleccionaron manualmente utilizando el paquete informático XMIPP (Sorzano *et al.*, 2004), para ser analizadas por el método de partículas individuales. El criterio de selección se basó en la coherencia en el tamaño mayoritario y la homogeneidad de la tinción sobre las partículas. Una vez seleccionadas, las partículas fueron extraídas de las micrografías en imágenes individuales de 140 x 140 píxeles y normalizadas para optimizar los resultados de la clasificación bidimensional (apartado 3.7.2). Después de su extracción las imágenes individuales fueron revisadas una segunda vez para descartar aquellas que presentaban zonas demasiado oscuras o claras que pudieran interferir en la clasificación. Finalmente se comenzó el procesamiento con un total de 13.273 partículas que mostraban un tamaño consistente y una tinción similar (Figura 14). En primer lugar se analizó la muestra mediante la clasificación bidimensional de las partículas utilizando los paquetes informáticos XMIPP (Scheres *et al.*, 2005a; Scheres *et al.*, 2005b) y EMAN (Ludtke *et al.*, 1999). El resultado de la clasificación con ambos programas mostró una forma mayoritaria del complejo similar a la identificada en las micrografías y diversas imágenes correspondientes a distintas proyecciones de la misma estructura. El tamaño del complejo en la vista mayoritaria es de ± 220 Å de largo por ± 115 Å de alto (Figura 14).

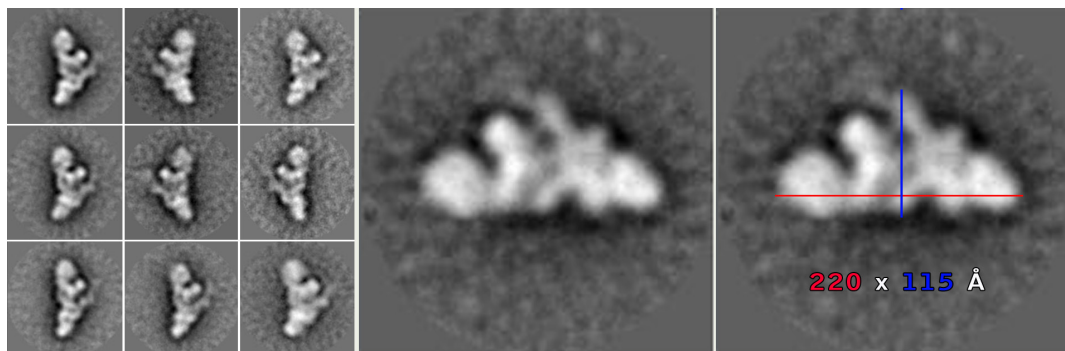


Figura 14: **Análisis bidimensional del complejo THO:Tex1 de *S. cerevisiae*.** La imagen muestra los resultados de realizar una clasificación bidimensional con nueve referencias (izquierda) y la media de la proyección mayoritaria del complejo con sus respectivas medidas de ancho y largo (derecha). En la figura se aprecia que el complejo presenta una forma característica en la que sus componentes se organizan a lo largo de un eje principal que diferencia una cara mas plana de otra mas abrupta. En la cara mas abrupta se distinguen dos salientes que destacan sobre el eje principal del complejo. El complejo tiene unas dimensiones de ± 220 Å de largo por ± 115 Å de alto.

A continuación se procedió a reconstruir tridimensionalmente el complejo THO:Tex1. Como se explicó en el apartado 3.7.3, para comenzar el refinamiento angular iterativo es necesario un volumen de referencia o modelo inicial. Al ser la primera vez que se analizaba estructuralmente este complejo no existían datos previos para ser utilizados como volumen de referencia, por lo que se procedió a generar computacionalmente volúmenes de referencia por los procedimientos de líneas comunes, *blob* y ruido (apartado 3.7.3.1). Con el objetivo de garantizar un resultado fiable, el refinamiento angular iterativo se realizó varias veces en paralelo partiendo de diferentes volúmenes de referencia. De esta forma se comprobó que todos los refinamientos conducían al mismo resultado independientemente del volumen de referencia empleado. El refinamiento angular iterativo evolucionó de forma uniforme y la estructura general del complejo se pudo adivinar a desde los primeros ciclos del procesamiento (Figura 15).

La evolución del procesamiento se realizó en base a un criterio de mantenimiento de la coherencia entre las proyecciones experimentales del complejo y los valores angulares asignados a cada proyección. En este punto fue de gran utilidad la representación angular de proyecciones implementada en el paquete informático EMAN, que esquematiza la distribución angular de partículas en un triángulo que resume los ángulos y las partículas asignadas a cada proyección teórica (Figura 15).

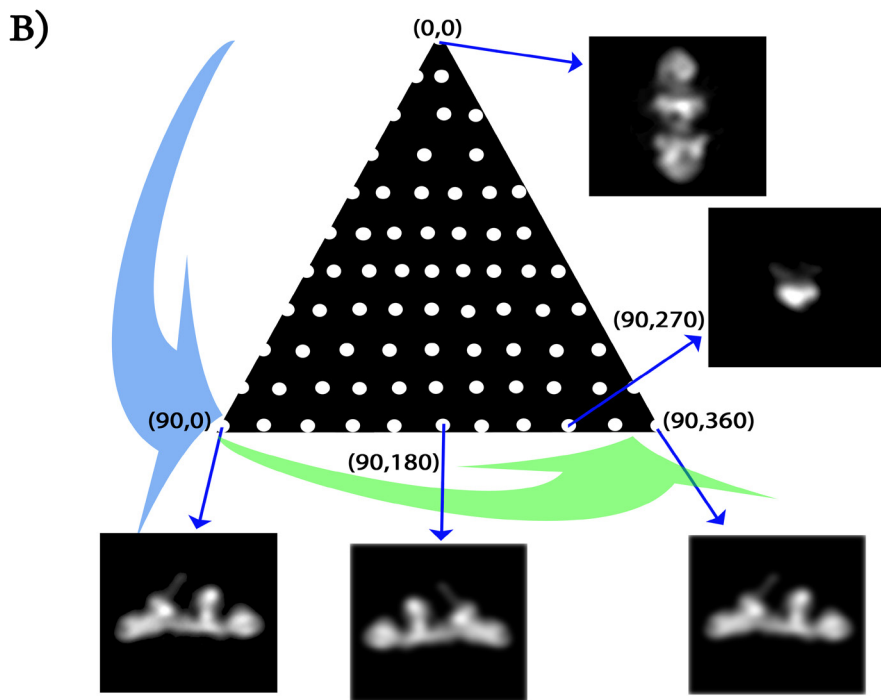
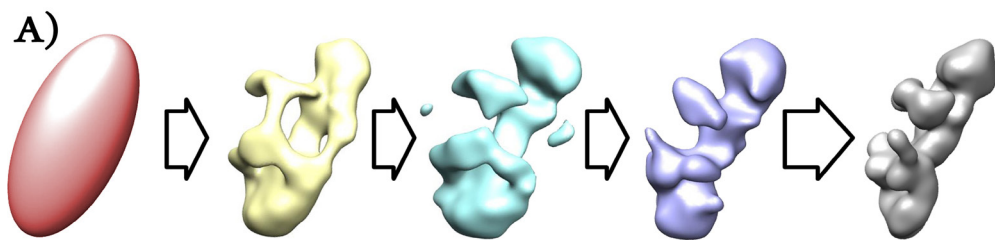


Figura 15: Evolución del refinamiento en la reconstrucción del complejo THO:Tex1. A) Resumen de la evolución del proceso de refinamiento angular iterativo a partir de un volumen de referencia de tipo *blob* (en rojo, dimensiones 60 x 40 x 20 píxeles volumétricos) a través de sucesivos ciclos de refinamiento angular iterativo. Se aprecia como con la estructura tridimensional del complejo se va definiendo a lo largo del proceso. Cada flecha representa varios ciclos de refinamiento angular iterativo realizados con EMAN y la flecha final, mas larga, incluye el procesamiento con otros paquetes informáticos (XMIPP y SPIDER) hasta alcanzar el volumen final (en gris). El criterio principal para avanzar en el refinamiento fue el mantenimiento de una distribución angular de partículas uniforme y consistente según el triángulo de proyecciones de EMAN. **B)** Esquema teórico simplificado del triángulo de proyecciones de EMAN en el que se representa la asignación de ángulos y partículas utilizada en la reconstrucción del complejo THO:Tex1. En la figura, la flecha azul muestra la distribución de las proyecciones en el triángulo según su valor de tilt, mientras que la flecha verde muestra la distribución según su valor de rot. Para explicar la distribución de proyecciones en el triángulo se han representado algunas proyecciones que responden a una distribución coherente. La primera proyección (arriba) muestra una vista apical del complejo, asignada a los ángulos de proyección (0,0). La esquina inferior izquierda del triángulo muestra una vista lateral asignada a los ángulos (90,0) mientras que la esquina inferior derecha muestra una vista muy similar asignada a los ángulos (90,360). En la zona inferior central del triángulo, se muestra una proyección frontal del complejo asignada a los ángulos (90,180). Esta distribución de proyecciones es consistente dado que la proyección asignada a los valores angulares (90,0) es muy similar a la asignada a los valores angulares (90,180) y (90,360), mientras que la asignada a los valores (90,270) responde a una proyección en el eje perpendicular a las anteriores. La visión apical asignada a los ángulos (0,0) es diferente a todas las registradas en la base del triángulo y también es consistente con las proyecciones asignadas a los valores angulares (90,0), (90,180), (90,360) y (90,270).

En el triángulo de proyecciones de EMAN cada proyección está definida por dos valores angulares, el primero indica la inclinación en el eje vertical (*tilt*) y el segundo en el eje horizontal (*rot*), ambos expresados en grados. Los valores angulares se representan entre paréntesis de la siguiente forma: (*tilt*, *rot*). Al tratarse de proyecciones el valor angular *tilt* oscila entre los valores 0° y 90° mientras que el valor *rot* lo hace entre los valores 0° y 360° . La combinación de valores para estos dos valores angulares puede generar cualquier proyección posible en el espacio tridimensional. En el triángulo, cada proyección está representada por un punto blanco de forma que el vértice superior representa las proyecciones generadas con los valores angulares (0,0) mientras que en la base del triángulo se representan en el vértice izquierdo las proyecciones en (90,0) y en el derecho las proyecciones en (90,360).

Los distintos volúmenes de referencia se refinaron de forma independiente hasta obtener un resultado convergente. El hecho de que los refinamientos angulares iterativos mostraran resultados similares con independencia del modelo inicial valida el resultado obtenido y demuestra que el volumen de referencia inicial no es determinante para la obtención de la estructura final del complejo THO:Tex1. Sin embargo, a lo largo del procesamiento se identificó una pequeña región flexible localizada en la parte superior del complejo (lado convexo) que presentaba cierta variabilidad estructural y mostraba leves diferencias entre las reconstrucciones procedentes de distintos refinamientos. Dado que el método de reconstrucción tridimensional de partículas individuales se basa en la obtención de medias bidimensionales que optimizan la relación señal ruido, la caracterización de regiones flexibles es un objetivo particularmente difícil para el análisis de muestras por microscopía electrónica. Así, las zonas flexibles presentan variabilidad entre las diferentes partículas del conjunto de imágenes y no quedan definidas en las medias bidimensionales, lo que finalmente provoca mayor variabilidad en las reconstrucciones tridimensionales.

El resultado de la reconstrucción tridimensional de complejo THO:Tex1 mostró una estructura muy característica, alargada y organizada entorno a un eje principal que diferencia una cara más abrupta de otra más lisa. Además, se observó que las proyecciones generadas en el último paso del refinamiento angular iterativo son compatibles con las imágenes medias obtenidas por los algoritmos de clasi-

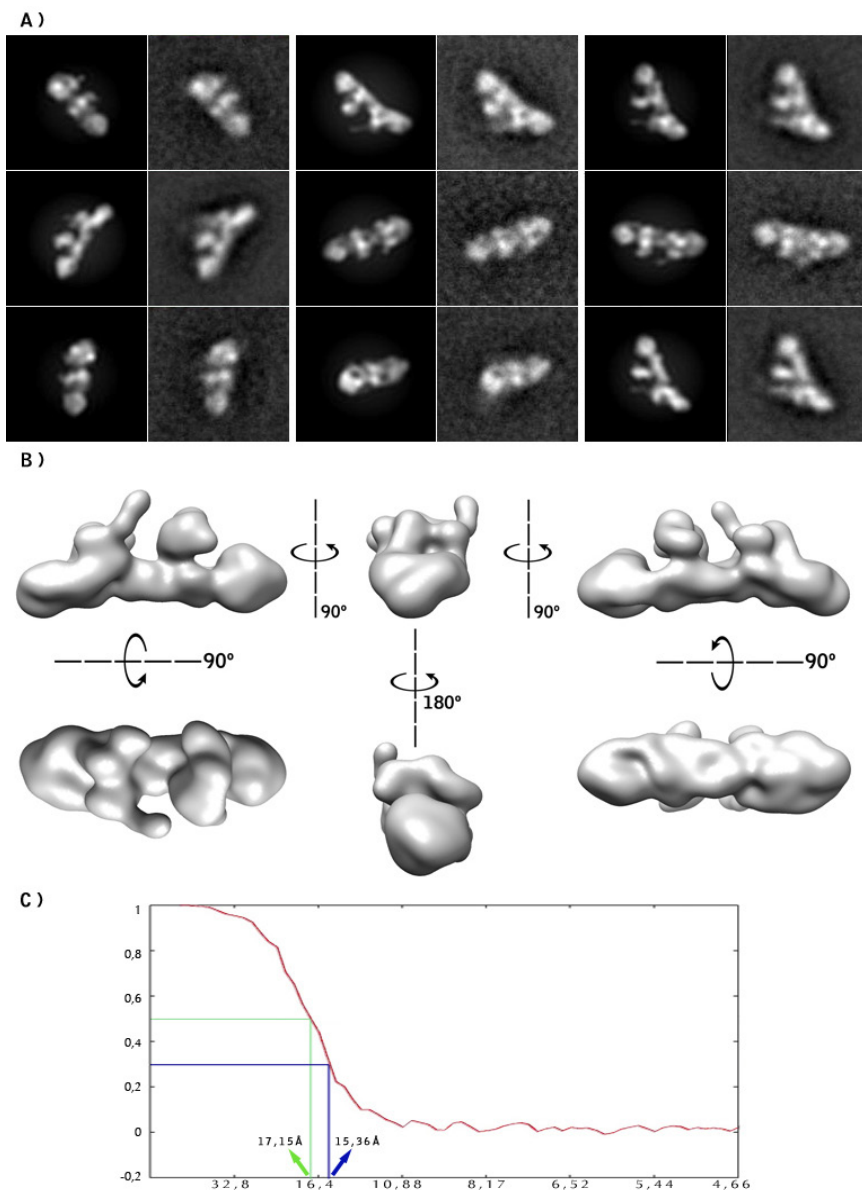


Figura 16: **Reconstrucción tridimensional del complejo THO:Tex1.** **A)** Asignación de partículas. La imagen muestra la asignación de partículas (proyecciones experimentales) a las proyecciones del volumen de referencia (proyecciones teóricas) en el último ciclo del refinamiento angular iterativo. La figura está organizada por pares de imágenes, de forma que se muestra primero la proyección teórica y luego la media de las partículas asignadas a esa proyección. Se observa que la imagen promedio de las partículas asignadas mantiene un fuerte parecido con la proyección del volumen de referencia, lo que sugiere que en el último ciclo de refinamiento el volumen de referencia es bastante similar al correspondiente a las proyecciones experimentales, es decir, al espécimen real de que proceden las partículas. **B)** Reconstrucción final. La figura muestra seis vistas diferentes de la reconstrucción final del complejo THO:Tex1. Las distintas vistas están organizadas en función de las flechas que indican el ángulo de giro correspondiente a cada vista. Se observa que la reconstrucción tridimensional presenta una estructura alargada muy característica en la cual se distingue una superficie relativamente plana frente a otras mas abrupta con dos salientes que destacan sobre el plano principal del complejo. **C)** Cálculo de la resolución. La imagen muestra la curva de la FSC en la que se han trazado líneas horizontales a la altura del valor 0,5 (línea verde) y de 0,3 (línea azul) para estimar la resolución según ambos criterios (apartado 3.7.3.3). La resolución de la estructura está en torno a los 16 Å.

cación bidimensional de EMAN y XMIPP, lo que sugiere que el volumen resultante es similar al volumen original a partir del cual se generaron las proyecciones experimentales (partículas) (Figura 16). Una vez finalizado el proceso de reconstrucción se llevó a cabo la estimación de la resolución de la estructura tridimensional resuelta atendiendo a la gráfica de la FSC (apartado 3.7.3.3). La resolución final de la estructura está entorno a los 16 Å (Figura 16).

En condiciones ideales, para reconstruir tridimensionalmente un espécimen habría que disponer de proyecciones experimentales del volumen en todas las direcciones del espacio. Es decir, si se dispone de un número limitado de proyecciones del espécimen la estimación de la tridimensionalidad puede no ser del todo precisa. Para aumentar el número de ángulos de proyección obtenidos a partir del complejo THO:Tex1 las micrografías se tomaron variando el ángulo de proyección de la rejilla en el microscopio electrónico, como se explicó en el apartado 3.6.2. Con el objetivo de comprobar que el conjunto de partículas seleccionadas abarcaba un rango amplio de proyecciones del modelo final, se atendió a los valores angulares asignados a cada partícula y al número de partículas incluidas en cada ángulo de proyección. Con estos datos se procedió a generar representaciones en dos y tres dimensiones, compuestas por esferas que se distribuyen entorno a la estructura resuelta, de forma que cada esfera representa un punto de proyección definido por unos valores angulares y su tamaño es proporcional a la cantidad de partículas asignadas a ese determinado punto de proyección (cantidad de proyecciones experimentales asignadas a las proyecciones teóricas) en el último ciclo de refinamiento angular iterativo (Figura 17).

A pesar de la forma alargada del complejo, el resultado mostró una distribución angular de partículas que cubría la mayoría de los ángulos de proyección, habiendo mayor cantidad de partículas asignadas a proyecciones laterales, en las que el complejo se sitúa con su eje mayor en paralelo a la superficie de la rejilla. Esta observación es además consistente con los datos observados en el análisis bidimensional del complejo (Figura 14).

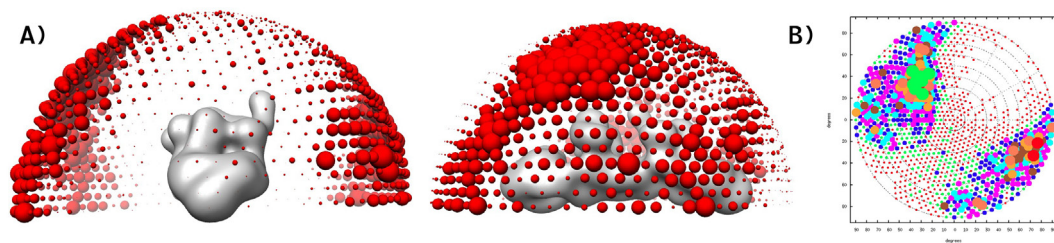


Figura 17: **Distribución angular de partículas en la reconstrucción del complejo THO:Tex1.** **A)** El esquema representa la distribución angular de partículas en el último ciclo del refinamiento angular iterativo de la reconstrucción del complejo THO:Tex1. En la figura la posición de cada esfera (color rojo) representa un punto de proyección (definido por sus valores angulares) y su tamaño es proporcional a la cantidad de partículas asignadas a la proyección teórica correspondiente a esos valores angulares. Se observa que, de acuerdo con la forma alargada del complejo, la mayoría de las partículas representan proyecciones laterales del complejo, mientras que las proyecciones en las que el eje mayor del complejo queda perpendicular a la superficie de la rejilla, son las menos abundantes. **B)** Representación similar a la anterior resumida en dos dimensiones. La figura representa la distribución angular de partículas en el último ciclo del refinamiento angular iterativo del complejo THO:Tex1 de forma que cada circunferencia representa un punto de proyección (definido por sus valores angulares) y su tamaño es proporcional al número de partículas asignado a la proyección teórica correspondiente a esos valores angulares. El esquema de color de las circunferencias responde a un patrón que permite diferenciar unas circunferencias de otras cuando estas se solapan y no aporta ninguna información adicional al esquema. Al igual que en la representación anterior se observa que la mayoría de las partículas han sido asignadas a proyecciones laterales del complejo.

4.4. Purificación del complejo THO a partir de la cepa THO Δ Tex1

Nuestro siguiente objetivo fue aislar el complejo THO de *S. cerevisiae* sin la proteína Tex1, tal y como fue descrito inicialmente (Chávez *et al.*, 2000). Los datos bibliográficos previos señalaban que las cepas mutantes en la proteína Tex1 no presentaban hiper-recombinación mitótica ni defectos en la transcripción y que la eliminación de Tex1 en *S. cerevisiae* no mostraba ningún efecto sobre en ruta de expresión génica (Luna *et al.*, 2005; García-Rubio *et al.*, 2008). En base a estos datos generamos la cepa mutante THO Δ Tex1 (apartado 3.1.3). La cepa THO Δ Tex1 presentaba un crecimiento similar al observado para la cepa original y dado que también contenía una cola TAP en el extremo carboxi-terminal de la proteína Tho2, el protocolo de purificación se realizó siguiendo los mismos pasos descritos para la cepa original Tho2-TAP (apartado 3.1.2). El resultado de la purificación fue el esperado: el gel SDS-PAGE mostraba las cuatro bandas correspondientes a las cuatro subunidades que componen el complejo THO de *S. cerevisiae*: Tho2, Hpr1, Mft1 y Thp2 (Figura 18).

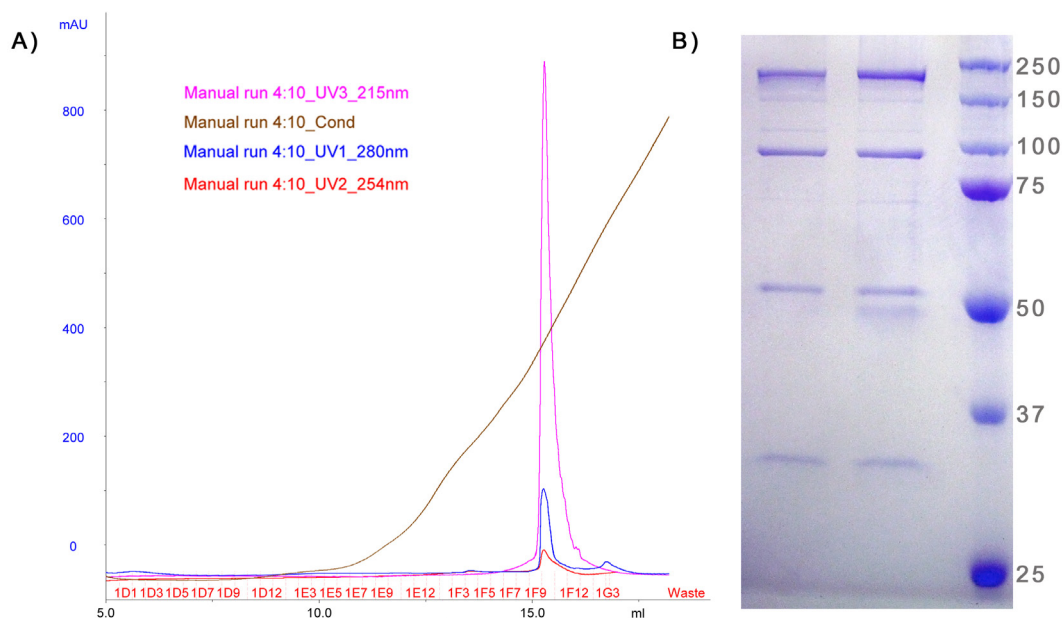


Figura 18: **Purificación del complejo THO a partir de la cepa THOΔTex1.** A) La imagen muestra el resultado de la cromatografía de intercambio iónico en una columna *Resource Q* realizada como paso final de la purificación en tándem por afinidad a partir de la cepa THOΔTex1. En el gráfico se presenta la absorción a 215, 280 y 254 nanómetros de longitud de onda (colores rosa, azul y rojo respectivamente) y la conductividad (color marrón), que aumenta a lo largo de la cromatografía con la creciente fuerza iónica en el interior de la columna. El pico del cromatograma comprendido entre las fracciones 1F8 y 1C11 corresponde al complejo THO (primer carril del gel presentado en la figura B). La concentración salina del tampón en el momento de la elución del complejo es de aproximadamente 400 mM, en un tampón 10mM Tris-HCl pH 8.0. B) La figura muestra un gel SDS-PAGE teñido con colorante azul de Coomassie. En el primer carril se cargaron aproximadamente 40 μl de la muestra resultante de las fracciones 1F8 a 1F11 del cromatograma presentado en la figura A y en el segundo carril se cargaron 40μl de la muestra resultante de la purificación a partir de la cepa Tho2-TAP (figura 11). Se aprecia que la primera muestra presenta una banda más que la segunda, entorno a un tamaño de 50 kDa aproximadamente: la proteína Tex1. Las otras bandas corresponden a las cuatro proteínas del complejo THO: Tho2, Hpr1, Mft1 y Thp2 (en orden de tamaño). Se ha marcado con números grises el tamaño correspondiente a cada marcador de peso molecular (MwM).

Los complejos THO resultantes de la purificación a partir de la cepa THOΔTex1 se degradaban mas rápidamente que los complejos THO:Tex1 obtenidos a partir de la cepa Tho2-TAP. Las muestras de complejo THO dejaban de ser utilizables después de dos días a partir de su purificación. Esta observación sugiere que una vez purificados, los complejos THO:Tex1 son mas estables que los complejos THO.

4.5. Cálculo estequiométrico de los componentes del complejo THO

A pesar de que el complejo THO de *S. cerevisiae* está compuesto por cuatro proteínas, existen estudios que proponen que el complejo contiene dos copias de cada uno de sus componentes (Oppizzi, 2004) y otros que muestran ensayos de cromatografía de exclusión molecular en los que los componentes del complejo THO parecen formar parte de un complejo de tamaño superior a los 500 kDa (Chang *et al.*, 1999). Por el contrario, otros autores sugieren que las subunidades del complejo se agrupan *in vivo* en relación estequiométrica 1:1:1:1 (Chávez *et al.*, 2000; Strasser *et al.*, 2002; Rehwinkel *et al.*, 2004). Con el objetivo de profundizar en esta cuestión se procedió a realizar una aproximación analizando el patrón de densidad de las bandas de cada proteína según su registro en un gel SDS-PAGE al 10% de acrilamida teñido uniformemente con colorante azul de Coomassie. Para aumentar la fiabilidad de los resultados, la medición del patrón de densidad se realizó en varios geles diferentes y con muestras de diferentes purificaciones. El resultado mostró una relación aproximadamente equimolar entre los componentes del complejo (Figura 19).

RESULTADOS

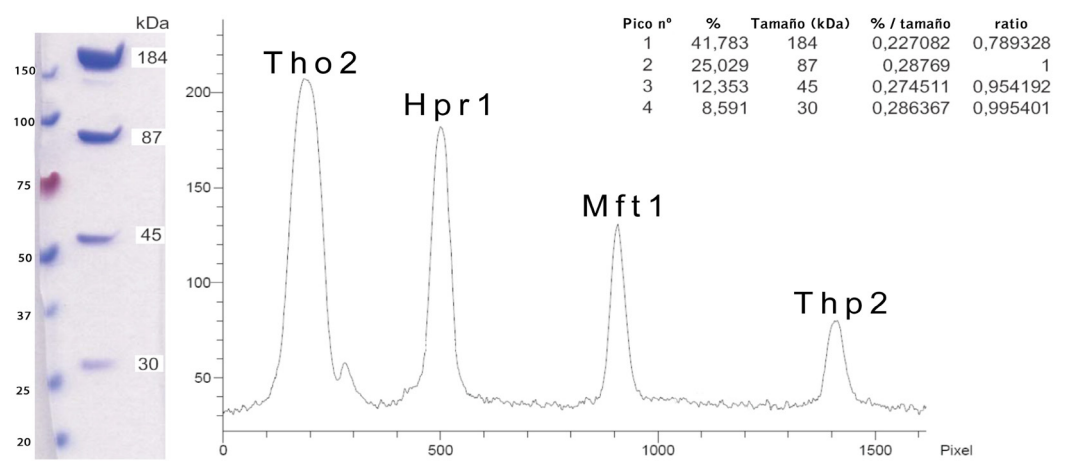


Figura 19: **Análisis del patrón de densidad de las subunidades del complejo THO.** La figura muestra un gel SDS-PAGE teñido con colorante azul de Coomassie (izquierda) y el patrón de absorbanza de cada una de sus bandas (derecha). A la izquierda del gel se han representado los tamaños correspondientes a cada banda de los marcadores de peso molecular (en kDa) y a la derecha el tamaño aproximado de cada una de las bandas registradas. En la esquina superior derecha de la figura se muestra la tabla con los valores numéricos de absorbanza para cada banda del complejo THO, tamaño de cada proteína y el porcentaje que representa con respecto a las demás. Los datos sugieren una relación equimolar entre los componentes del complejo THO de *S. cerevisiae*. Teniendo en cuenta estos resultados y las dimensiones resultantes de la observación del complejo por microscopía electrónica, se propone que el complejo THO de *S. cerevisiae*, en su forma nativa, está compuesto por una de cada una de las cuatro proteínas que lo componen.

4.6. Reconstrucción tridimensional del complejo THO

Para el procesamiento del complejo THO se utilizó la muestra resultante de la purificación a partir de la cepa mutante THO Δ Tex1. Las rejillas de microscopía electrónica se tiñeron con acetato de uranilo al 2% (apartado 3.6.1) y se observaron en el microscopio electrónico de transmisión. Las micrografías registradas mostraban una distribución homogénea de partículas con una forma y tamaño constante. La observación a través de los binoculares del microscopio permitió distinguir una estructura similar a la observada en el complejo THO:Tex1. Se tomaron micrografías y se seleccionaron manualmente 14.115 partículas correspondientes a diferentes proyecciones del complejo (Figura 20). Al igual que en el procesamiento anterior, el criterio de selección de partículas se basó en su tamaño, la tinción y la calidad de las mismas. Las partículas se extrajeron, normalizaron y agruparon en un conjunto de imágenes de 140 x 140 píxeles. Para el análisis bidimensional las partículas se redimensionaron a las mitad de su tamaño con el objetivo de optimizar los tiempos de cálculo computacional. El procesamiento bidimensional se llevó a cabo, al igual que en el caso del complejo THO:Tex1, utilizando los paquetes informáticos de EMAN y XMIPP. El resultado de ambos procesamientos mostró una proyección mayoritaria y varias vistas de la misma estructura desde diferentes ángulos de proyección (Figura 20).

A continuación se procedió al análisis tridimensional del complejo THO. Para ello se utilizaron cuatro modelos iniciales diferentes que se refinaron en paralelo, comprobando que la evolución de todos ellos avanzaba hacia un volumen tridimensional similar. Los modelos iniciales se generaron a partir líneas comunes, *blob*, ruido y del modelo de THO:Tex1 filtrado a una resolución de a 42 Å. Utilizando parámetros similares a los empleados en la reconstrucción del complejo THO:Tex1 se comprobó que los volúmenes resultantes de los diferentes refinamientos presentaban una estructura similar a la obtenida para el complejo THO asociado a la proteína Tex1. Los componentes del complejo THO parecían distribuirse a lo largo de un eje principal que confiere una estructura alargada al complejo y separa una cara mas abrupta de otra mas lisa (Figura 21). Las proyecciones del volumen de referencia

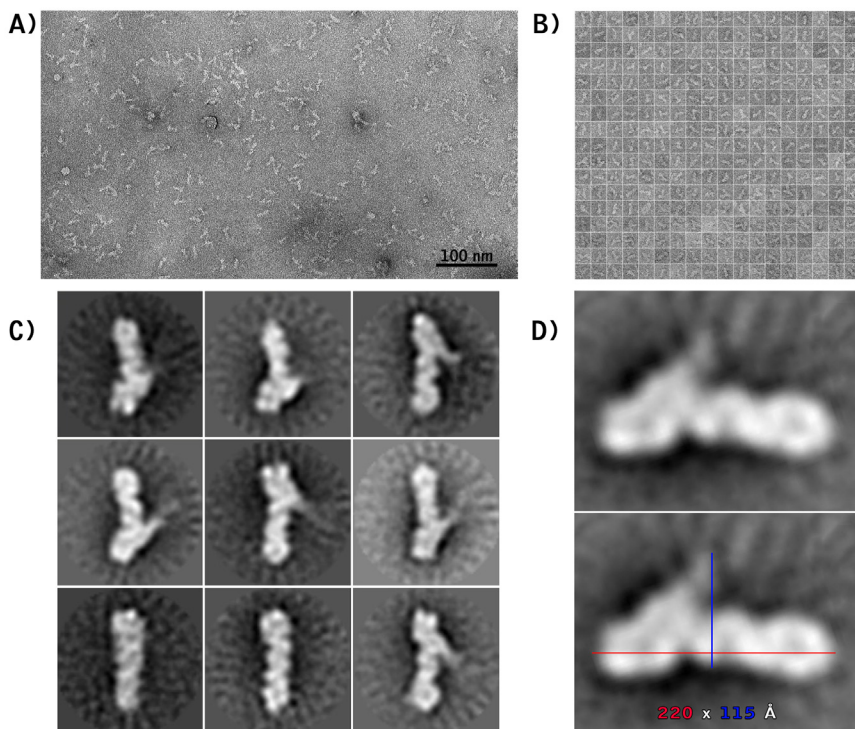


Figura 20: Procesamiento bidimensional del complejo THO. **A)** Micrografía. La imagen muestra una micrografía de tinción negativa obtenida a partir de una rejilla preparada con la muestra resultante de la purificación a partir de la cepa THO Δ Tex1. La imagen se tomó a 60.000 aumentos (barra de escala en la esquina inferior derecha). Se aprecian varias vistas de una estructura que se repiten de forma mas o menos uniforme en el campo de visión: el complejo THO. **B)** Conjunto de imágenes. Vista parcial del conjunto de partículas seleccionadas manualmente y cortadas de las micrografías en imágenes de 140 x 140 píxeles. **C)** Resultado de la clasificación bidimensional del conjunto de partículas en nueve clases. Las medias bidimensionales revelan diferentes proyecciones del complejo. **D)** Imagen promedio que muestra la proyección mayoritaria del complejo, con su eje mayor situado en paralelo al plano de la superficie de la rejilla. Las dimensiones aproximadas del complejo se han representado por medio de las líneas roja (largo 220 Å) y azul (alto 115 Å). Las dimensiones del complejo coinciden con las observadas en el complejo THO:Tex1 (Figura 14).

generadas en el último paso del refinamiento angular iterativo (proyecciones teóricas) eran muy similares a las imágenes medias obtenidas en la clasificación bidimensional y se ajustaban bien a las medias de partículas asignadas a cada proyección (Figura 21). Esta observación confirma que el volumen de referencia utilizado en la última vuelta del refinamiento angular iterativo es muy parecido al volumen original del que provienen las partículas, es decir al espécimen. La coherencia y similitud entre las partículas y las proyecciones del volumen final contribuye a la validación de la estructura resuelta. Una vez finalizada la reconstrucción tridimensional se llevó a cabo la estimación de la resolución de la estructura resultante atendiendo a la gráfica de la FSC (apartado 3.7.3.3). La resolución final de la estructura está entorno a los 19 Å (Figura 21).

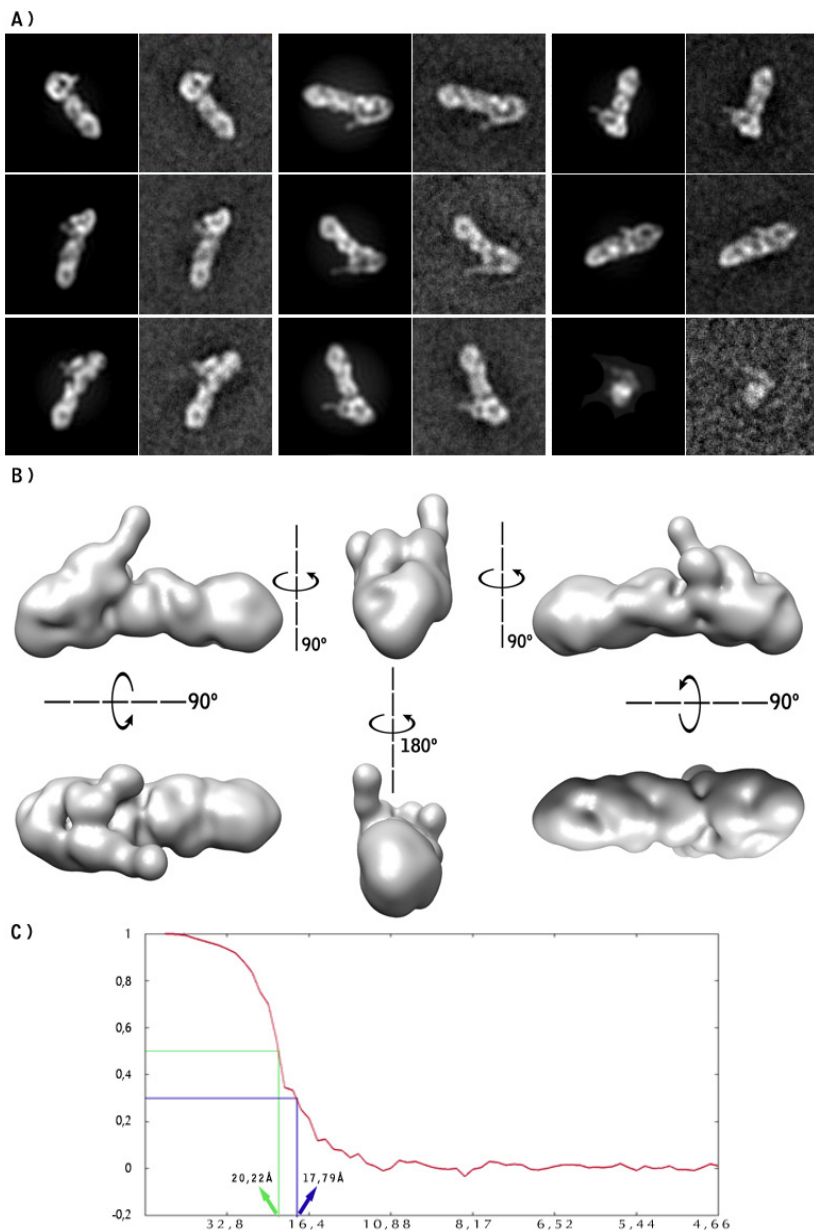


Figura 21: Reconstrucción tridimensional del complejo THO. **A)** Asignación de partículas. La imagen muestra la asignación de partículas (proyecciones experimentales) a las proyecciones del volumen de referencia (proyecciones teóricas) en el último ciclo de refinamiento angular iterativo de la reconstrucción del complejo THO. La figura está organizada por pares de imágenes, de forma que se muestra primero la proyección teórica y luego la media de las partículas asignadas a esa proyección. Se observa que la imagen promedio de las partículas asignadas guarda un alto parecido con la proyección del volumen de referencia, lo que sugiere que en el último ciclo de refinamiento el volumen de referencia es bastante similar al correspondiente a las proyecciones experimentales, es decir, al volumen real de que proceden las partículas (especimen). **B)** Reconstrucción final. La figura muestra seis vistas diferentes de la reconstrucción final del complejo THO. Las distintas vistas están organizadas en función de las flechas que indican el ángulo de giro correspondiente a cada vista. Se observa que la estructura se organiza en función de un eje mayoritario que divide el complejo en una cara mas plana y otra mas abrupta. **C)** Cálculo de la resolución. La imagen muestra la curva de la FSC en la que se han trazado líneas horizontales a la altura del valor 0,5 (línea verde) y de 0,3 (línea azul) para estimar la resolución según ambos criterios (apartado 3.7.3.3). La resolución de la estructura está en torno a los 19 Å.

4.7. Predicción de estructura de la proteína Tex1

En base a estudios previos que presentan a la proteína Tex1 de *S. cerevisiae* como miembro de la familia de las proteínas WD (Rehwinkel *et al.*, 2004), se decidió generar un modelo atómico computacional de esta proteína, basado en la comparación de su secuencia con la de otras proteínas de estructura conocida (threading).

La proteína Tex1 presenta repeticiones de tipo WD40, que es un motivo estructural pequeño (aproximadamente 40 aminoácidos) terminando típicamente en los aminoácidos triptófano y ácido aspártico (representados por las letras W y D, respectivamente). Frecuentemente estas repeticiones se encuentran incluidas en un plegamiento tridimensional característico conocido como β -propeller (Rehwinkel *et al.*, 2004) que consiste en una distribución de la secuencia de aminoácidos en forma de láminas β que se organizan en pliegos dispuestos alrededor de un eje central en forma toroidal. Cada una de las hojas β generalmente presenta tres o cuatro láminas β antiparalelas dobladas de tal forma que la primera y la cuarta láminas son prácticamente perpendiculares entre si (Neer *et al.*, 1994; Smith *et al.*, 1999). Las proteínas WD se encuentran ampliamente representadas en todos los organismos eucariotas y funcionan frecuentemente facilitando las interacciones proteína-proteína (Fulop *et al.*, 1999; Smith *et al.*, 1999; Springer, 2002).

Para obtener un modelo atómico de la proteína Tex1 de *S. cerevisiae* se realizó, en primer lugar, una búsqueda de posibles homólogos con el programa BLAST (Altschul *et al.*, 1990) a partir de la secuencia aminoacídica de Tex1 (número de acceso de SwissProt P53851). Así, se identificaron las secuencias que mayor identidad presentaban con la secuencia de Tex1 y con ellas se realizó un alineamiento múltiple utilizando el programa MUSCLE (Edgar, 2004b). A continuación se visualizó y editó el alineamiento múltiple con el programa JALVIEW (Waterhouse *et al.*, 2009), con el que se mejoraron los resultados del alineamiento según criterio experto (apartado 3.5). Este análisis de la secuencia permitió obtener una visión general de la distribución de motivos a lo largo de la secuencia, así como de las regiones más conservadas y los posibles dominios estructurales. El alineamiento múltiple y el ajuste manual permitió aumentar la fiabilidad de la predicción de estructura secundaria,

que se hizo a continuación con el paquete informático JPRED3 (Cole *et al.*, 2008).

El resultado de la predicción mostró que la mayoría de la secuencia se organiza en una estructura secundaria de tipo láminas β a intervalos regulares (Figura 22). Este resultado es consistente con los datos bibliográficos previos que identificaban a Tex1 con un plegamiento de tipo β -propeller (Rehwinkel *et al.*, 2004).

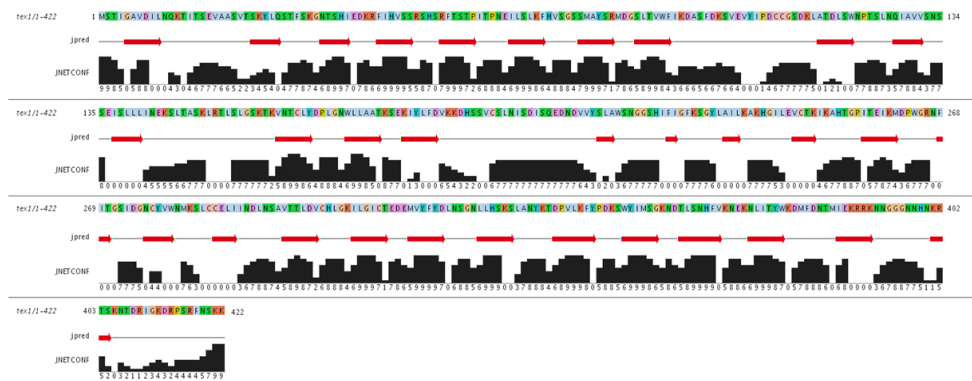


Figura 22: Predicción de la estructura secundaria de la proteína Tex1 de *S. cerevisiae*. La figura muestra el resultado simplificado de la predicción de estructura secundaria para la proteína Tex1. La secuencia de Tex1 (422 aminoácidos) aparece en la parte superior de cada fila, coloreada según el tipo de aminoácido. Los valores del parámetro “JPRED” corresponden al resultado final de la predicción (láminas β representadas mediante flechas rojas). Los valores del parámetro “JNETCONF” representan el valor de confianza de la predicción realizada, que oscila del 1 al 9 de forma que cuanto mayor es el valor mas fiable es la predicción (en color negro). En la figura se observa la secuencia de Tex1 presenta láminas β a intervalos mas o menos regulares. Los colores de los aminoácidos en la secuencia son los siguientes: en azul claro Alanina, Isoleucina, Leucina, Metionina, Fenilalanina, Valina y Triptófano (los mas apolares); en verde Treonina, Glutamina, Asparagina y Serina (los mas hidrofílicos); en rosa el Ácido Glutámico y el Ácido Aspártico (cargados negativamente); en naranja Arginina y Lisina (cargados positivamente); en turquesa Tirosina e Histidina; en amarillo Prolina; en rojo claro Cisteina; en naranja claro Glicina.

Con el objetivo de generar un modelo atómico de la estructura terciaria de Tex1, se introdujo su secuencia en las bases de datos de estructuras atómicas resueltas disponibles a través de los programas de predicción de estructura terciaria I-TASSER (Zhang, 2008; Roy *et al.*, 2010) y PHYRE (Kelley *et al.*, 2009). El resultado de ambos programas mostró modelos atómicos basados exclusivamente en proteínas con plegamientos de tipo β -propeller. Así, se seleccionaron aquellas estructuras que mostraban mejores valores de ajuste (candidatas): Aip1 (*actin interacting protein* de *C. elegans*, homóloga a la de levadura), Wr5 (*WD-repeat protein 5* de *H. sapiens*), Paf-Ah (*Platalet-acrivating factor acetylhidrolase* de *Mus musculus*), domino carboxi-terminal de Sif2 (*Sir4 interacting protein* de *S. cerevisiae*) y la subunidad β 1 del complejo transducina de *H. sapiens*. A continuación se utilizó Swiss-PDBviewer (Guex *et al.*, 1997; Guex *et al.*, 2009) para alinear la secuencia

de Tex1 sobre la estructura de cada una de las proteínas candidatas obtenidas. Los modelos obtenidos se evaluaron con las herramientas de validación de estructuras de SwissModel (Arnold *et al.*, 2006).

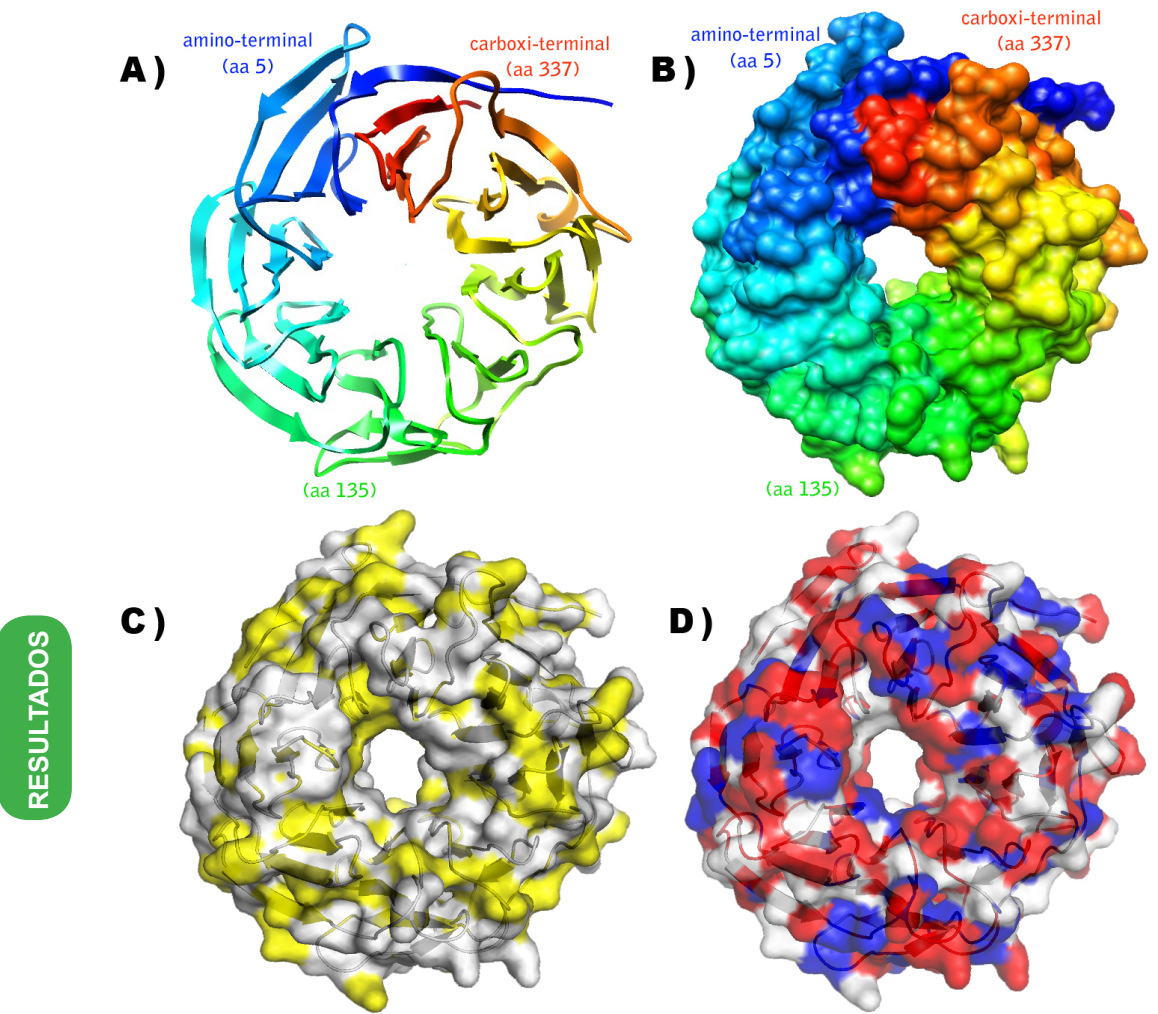


Figura 23: **Modelo atómico de la región amino-terminal de la proteína Tex1 de *S. cerevisiae*.** La figura muestra el modelo atómico de la proteína Tex1 generado computacionalmente en base a la comparación de su secuencia con la del dominio carboxi-terminal de la proteína Sif2 de *S. cerevisiae*. **A)** La figura muestra una representación simplificada de la secuencia coloreada de forma que comienza en color azul en el extremo amino-terminal (primer aminoácido incluido en el modelo es número 5 de la secuencia completa de Tex1) y avanza primero hacia abajo a la izquierda y luego hacia arriba a la derecha, hasta terminar en color rojo, que representa los aminoácidos mas próximos al extremo carboxi-terminal de la proteína (el aminoácido 337 es el último que aparece representado en el modelo). Se observa que la distribución de la secuencia responde a una estructura tridimensional de tipo β -propeller. **B)** Representación equivalente a la anterior en la que se muestra la superficie total de la molécula, manteniendo el mismo esquema de color que en La figura anterior. **C)** Representación similar a las anteriores en la que se han coloreado los aminoácidos apolares en amarillo y los polares en blanco. **D)** Representación similar a las anteriores en la que se han coloreado los aminoácidos con carga positiva en azul y los que tienen carga negativa en rojo. En las figuras C y D la superficie de la molécula se ha representado con un 40% de opacidad, lo que permite distinguir la estructura secundaria y la organización tridimensional de la secuencia en β -propeller.

El mejor resultado se obtuvo en el alineamiento de la secuencia de Tex1 con el extremo carboxi-terminal de la proteína Sif2 de *S. cerevisiae* (Cerna *et al.*, 2005). La secuencia aminoacídica de Tex1 muestra tan solo un 11% de identidad con la secuencia del extremo carboxi-terminal de Sif2 y a pesar de ello, ambas parecen presentar un plegamiento muy similar. Este fenómeno ocurre con bastante frecuencia en muchas proteínas y se conoce como convergencia evolutiva de estructura. La visualización del modelo atómico final se realizó con los programas Pymol y UCSF-Chimera (DeLano, 2002; Pettersen *et al.*, 2004; Goddard *et al.*, 2005) (Figura 23).

A pesar de que la secuencia original de la proteína Tex1 de *S. cerevisiae* consta de 422 aminoácidos, el modelo atómico obtenido comprende desde el aminoácido 5 hasta el 337, ya que no fue posible realizar la predicción de estructura terciaria para los 85 aminoácidos del extremo carboxi-terminal. Así, el modelo obtenido representa aproximadamente el 78% de la estructura completa de Tex1. Este hecho sugiere la presencia de dos dominios en la estructura de Tex1, un dominio amino-terminal organizado en un plegamiento de tipo β -propeller y otro dominio carboxi-terminal cuya estructura terciaria no ha podido ser modelada por homología con otras proteínas de estructura conocida. La posible existencia de dominios en la proteína Tex1 no ha sido descrita aún por otros autores.

4.8. Localización de la proteína Tex1 en el complejo THO:Tex1

Para determinar la localización de la proteína Tex1 en el complejo THO:Tex1 se procedió a comparar las proyecciones mayoritarias de los complejos obtenidos a partir de las cepas Tho2-TAP y THO Δ Tex1. De esta forma se observó simultáneamente el complejo THO en presencia y en ausencia de la proteína Tex1, lo que permitió identificar la posición de Tex1. Así se identificó que la proteína Tex1 interacciona con el complejo THO por su zona central, aproximadamente a la mitad de su eje longitudinal. Con estos datos se procedió a ajustar manualmente el modelo de la proteína Tex1 (región amino-terminal) en el volumen del complejo THO:Tex1. Los resultados mostraron que la masa identificada como Tex1 es compatible con el modelo en β -propeller de la región amino-terminal de la proteína Tex1 (Figura 24).

4.9. Determinación de la posición de Hpr1 en el complejo THO

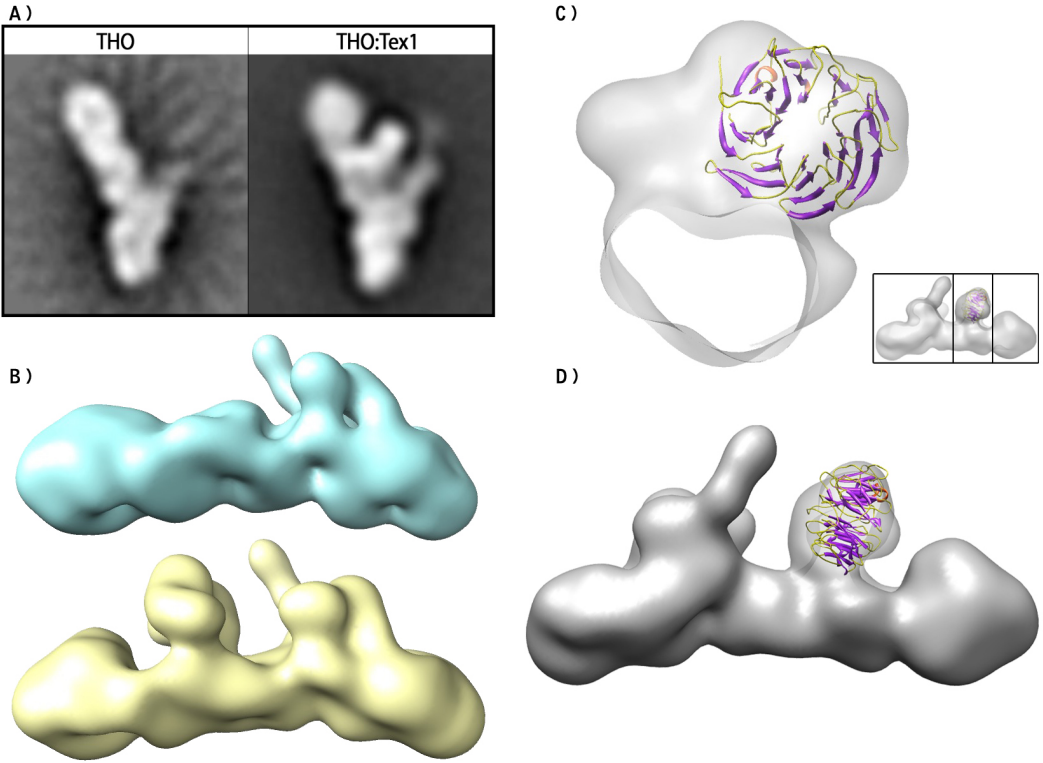


Figura 24: **Localización y ajuste del modelo de Tex1 en la estructura del complejo THO:Tex1.** **A)** La figura muestra las imágenes medias correspondientes a la proyección mayoritaria de los complejos THO y THO:Tex1. Se observa que la proyección lateral del complejo THO:Tex1 presenta una masa extra en comparación la del complejo THO. Esta masa se localiza en la zona central, aproximadamente a la mitad del eje longitudinal del complejo. **B)** La imagen muestra las reconstrucciones tridimensionales de los complejos THO (verde) y THO:Tex1 (amarillo). Se aprecia que entre ambas reconstrucciones son bastante similares excepto en la zona central donde el complejo THO:Tex1 presenta una masa extra ausente en el complejo THO. **C)** La figura muestra una vista frontal del ajuste del modelo atómico de la región amino-terminal de Tex1 en la sección del complejo THO:Tex1 correspondiente a esta proteína. Se observa que el modelo atómico de Tex1 (región amino-terminal) es compatible con la masa correspondiente identificada en el complejo THO:Tex1. Las secciones del corte se indican en el dibujo que aparece en la esquina inferior derecha de La figura, en un marco negro. Dado que el modelo atómico de Tex1 comprende aproximadamente el 78% de la proteína completa se podría proponer que el otro 22% restante estaría localizado en el lateral de la estructura en *β -propeller*, abarcando así el total o la mayoría de la masa correspondiente identificada. **D)** Ajuste del modelo atómico de Tex1 (región amino-terminal) en la estructura tridimensional del complejo THO:Tex1. En la imagen se aprecia que tanto el tamaño como la organización en *β -propeller* de la región amino-terminal de la proteína Tex1 son compatibles con el área correspondiente a esta proteína en la reconstrucción tridimensional realizada por microscopía electrónica.

A continuación se propuso identificar alguno de los componentes del complejo THO en la estructura tridimensional resuelta, con el objetivo de definir la topología de las subunidades en el complejo.

La inmunomicroscopía es una técnica que puede utilizarse para la identificación de subunidades en complejos macromoleculares por microscopía electrónica (Cuéllar *et al.*, 2008). Al no existir ningún anticuerpo comercial disponible para experimentos de microscopía, se pensó en diseñar uno, pero la falta de datos estructurales de las proteínas del complejo complicaba la tarea. Se intentó predecir la estructura atómica de las subunidades por separado, con el propósito de identificar que regiones pudieran quedar expuestas al medio y ser, por lo tanto, susceptibles de interaccionar con el anticuerpo. Sin embargo, ninguno de los componentes del complejo THO mostraba homología con proteínas de estructura conocida y tras consultar con expertos en la materia se llegó a la conclusión de que ésta opción no era viable.

A mediados del año 2005 se descubrió que la proteína Hpr1 es regulada por ubiquitinación durante el proceso de transcripción *in vivo* y se la propuso como responsable de la formación del complejo THO en *S. cerevisiae*. Gracias a la elaboración de mutantes en su extremo carboxi-terminal, se definió que los últimos ochenta y ocho aminoácidos constituyen el punto de regulación por ubiquitinación del complejo THO *in vivo* (Gwizdek *et al.*, 2005). Además, poco después se identificó al factor de exportación Mex67 como responsable de la estabilidad de Hpr1 y del complejo THO, por su capacidad de interaccionar con el dominio carboxi-terminal de Hpr1 y evitar así el marcaje por ubiquitinación (Gwizdek *et al.*, 2006). Además, datos previos apuntaban a que el complejo THO interacciona con la ARN-Pol II durante la transcripción a través de la proteína Hpr1 (Chang *et al.*, 1999) y que Hpr1 es a la vez responsable de la interacción con varios factores de exportación (Zenklusen *et al.*, 2002; Hurt *et al.*, 2004). Estos resultados sugieren que al menos el dominio carboxi-terminal de la proteína Hpr1 debe estar expuesta al medio durante la transcripción, en la forma nativa del complejo THO. Con estos datos se decidió diseñar un anticuerpo contra el extremo carboxi-terminal de Hpr1 (apartado 3.1.5).

El anticuerpo resultante se probó mediante *western-blot* y *dot-blot* para comprobar su interacción con el complejo de THO *S. cerevisiae*. En el *dot-blot*, el complejo THO:Tex1 recién purificado se aplicó directamente sobre una membrana de nitrocelulosa y se incubó con tampón de bloqueo antes de que la muestra se secase totalment. De esta forma los complejos quedaron solo parcialmente desnaturalizados sobre la membrana de nitrocelulosa (apartado 3.1.4). El resultado positivo en la interacción entre el anticuerpo anti-Hpr1 y el complejo THO en los ensayos de *dot-blot* supuso un indicio mas de que el epítipo del anticuerpo se mantenía expuesto al medio aún sin desnaturalizar y separar totalmente los componentes del complejo.

Una vez comprobada su especificidad, el anticuerpo anti-Hpr1 se incubó con la muestra del complejo THO:Tex1 (apartado 3.1.5). Debido a que ambos mostraban la misma capacidad de interacción con el anticuerpo anti-Hpr1 y a que una vez purificado el complejo THO:Tex1 es mas estable que el complejo THO en ausencia de Tex1 se decidió utilizar el primero para los ensayos de inmunolocalización. Después de la incubación la muestra se diluyó hasta una concentración final de entre 5 y 10 µg por mililitro y se prepararon varias rejillas de tinción negativa (apartado 3.2.1). Las muestras se observaron en un microscopio electrónico, se seleccionaron manualmente las partículas y se promediaron hasta obtener una media bidimensional de la vista mayoritaria del complejo THO:Tex1:antiHpr1 (Figura 25).

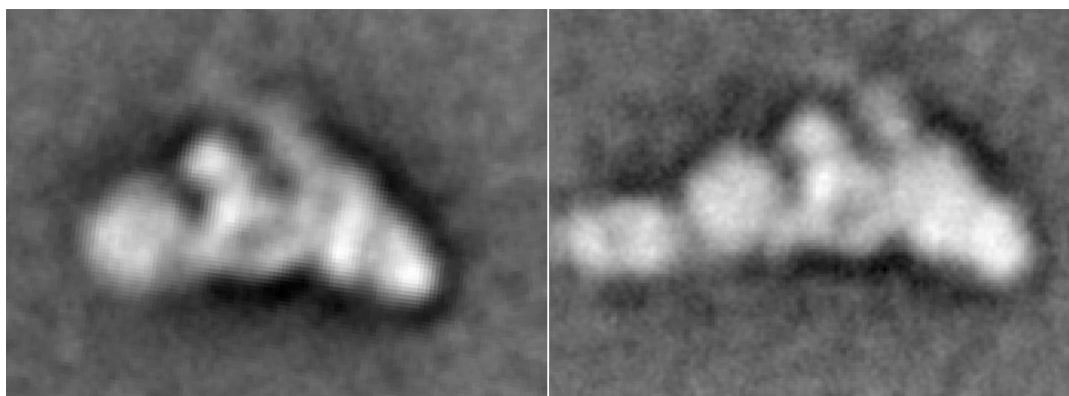


Figura 25: Identificación de la proteína Hpr1 en el complejo THO:Tex1. La figura muestra las medias bidimensionales de la proyección mayoritaria de los complejos THO:Tex1 y THO:Tex1:antiHpr1. Se aprecia que el complejo THO:Tex1:antiHpr1 muestra una masa consistente en la región posterior del complejo: el anticuerpo anti-Hpr1. El anticuerpo aparece representado sólo por una parte de su estructura total, dado que se trata de una molécula muy flexible que se ve parcialmente reducida de su tamaño real por el promediado de partículas utilizado durante la clasificación.

4.10. Caracterización de la interacción del complejo THO con el ADN y ARN

El complejo THO está implicado en la síntesis, metabolismo y exportación del ARNm en todos los organismos eucariotas desde levaduras hasta humanos (Stutz *et al.*, 2000; West *et al.*, 2000; Hammell *et al.*, 2002; Strasser *et al.*, 2002; Zenklusen *et al.*, 2002). El complejo funciona durante la elongación transcripcional y se sitúa junto a la ARNP_{II} que sintetiza el ARNm y cerca del ADN, al que parece proteger durante la transcripción (Jimeno *et al.*, 2002). Algunos estudios han propuesto que el complejo THO, en presencia de otros factores de exportación, es capaz de interaccionar directa o indirectamente con el ARNm y con el ADN (González-Barrera *et al.*, 2002; Jimeno *et al.*, 2002; Li *et al.*, 2005; Voynov *et al.*, 2006) pero su interacción con los ácidos nucleicos no ha sido aún caracterizada y no ha sido descrita ninguna subunidad implicada en dicha interacción.

Con el objetivo de estudiar la posible interacción del complejo THO con los ácidos nucleicos se incubó la muestra purificada a partir de la cepa Tho2-TAP con moléculas de ADN de cadena simple (ADNcs), ADN de cadena doble (ADNcd) y ARN marcadas radiactivamente (apartado 3.3). Después de la incubación la muestra fue expuesta a luz ultravioleta para inducir entrecruzamiento (*cross-link*) entre los nucleótidos y las posibles proteínas implicadas en la interacción. A continuación las muestras se sometieron a una electroforesis desnaturalizante en geles SDS-PAGE (apartado 3.1.4). El gel resultante se introdujo en una cámara de autorradiografía para detectar la presencia de los ácidos nucleicos marcados radiactivamente y a continuación se tiñó con colorante azul de Coomassie para marcar las proteínas e identificarlas en el gel. El resultado mostró que la única proteína del complejo THO involucrada en la interacción con los ácidos nucleicos ADN y ARN era Tho2, la proteína mas grande del complejo (Figura 26).

El siguiente objetivo fue comprobar si la proteína Tex1 influye en la interacción del complejo THO con los ácidos nucleicos. Para ello se repitió el mismo experimento utilizando las muestras obtenidas a partir de las cepas Tho2-TAP y THO Δ Tex1. El resultado mostró que la interacción entre la proteína Tho2 y los ácidos nucleicos (ADN, ARNcd y ADNcs) es similar en el caso de los complejos THO y THO:Tex1, lo que sugiere que Tex1 no influye en dicha interacción (Figura 26).

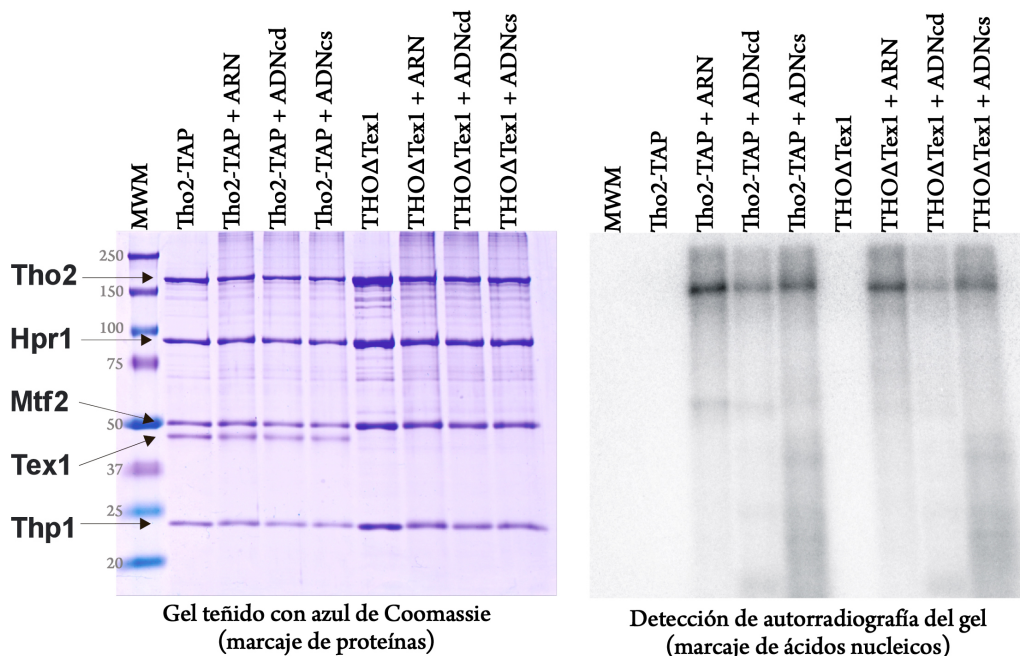
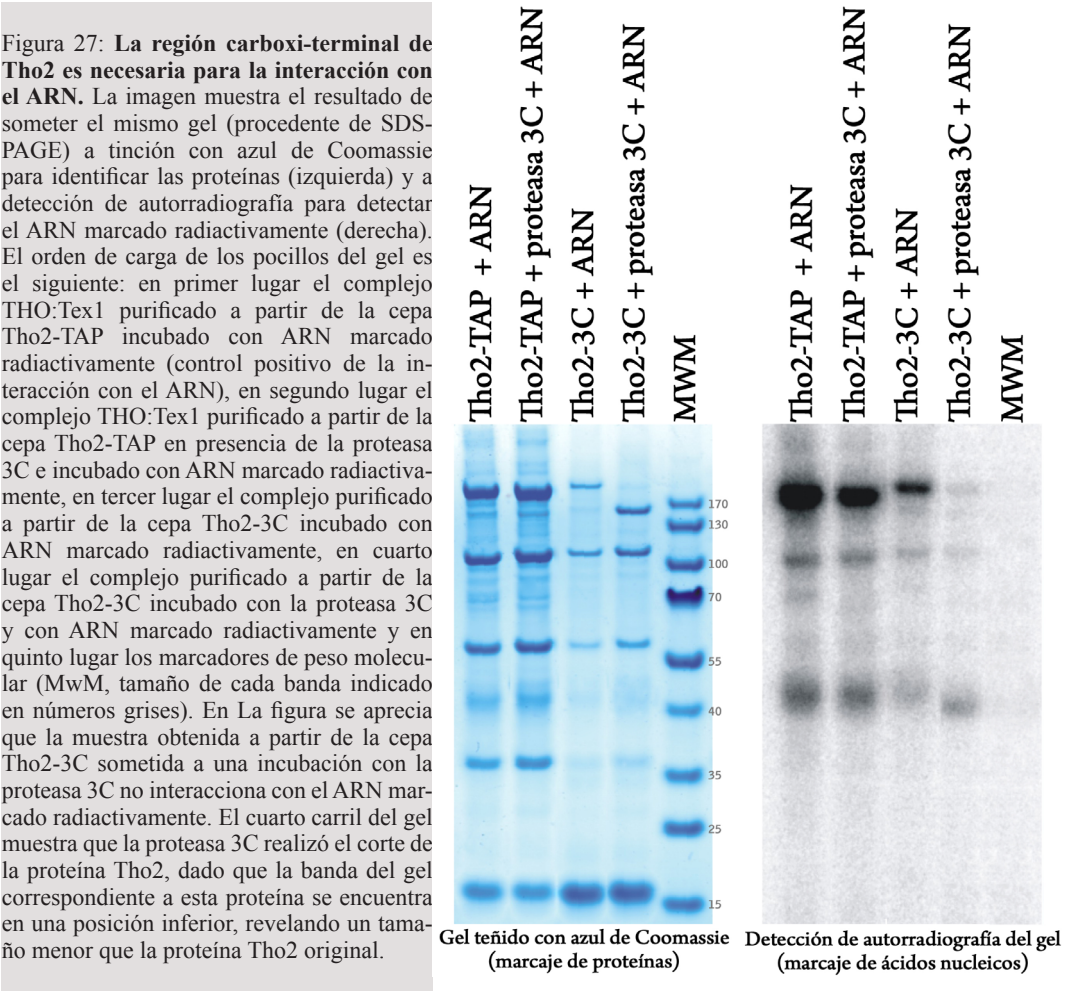


Figura 26: **Tho2 interactúa con los ácidos nucleicos en presencia y ausencia de Tex1.** La imagen muestra el resultado de someter el mismo gel (procedente de SDS-PAGE) a tinción con azul de Coomassie para marcar las proteínas (izquierda) y a detección de autorradiografía para detectar los ácidos nucleicos marcados radiactivamente (derecha). Se han utilizado muestras procedentes de las cepas de *S. cerevisiae* Tho2-TAP y THOΔTex1 para comprobar si la proteína Tex1 interviene en la interacción del complejo THO con los ácidos nucleicos. En la figura se aprecia que la única proteína del complejo que interactúa con los ácidos nucleicos es Tho2, la proteína mas grande del complejo y que lo hace de forma independiente de la presencia de Tex1. El orden de carga de los pocillos del gel es, en primer lugar el marcador de peso molecular (MwM, tamaño de cada banda indicado en números grises), en segundo el complejo THO:Tex1 obtenido a partir de la cepa Tho2-TAP (control negativo para el marcaje de autorradiografía), en tercer lugar el complejo THO:Tex1 incubado con ARN marcado radiactivamente, el cuarto el complejo THO:Tex1 incubado con ADN de cadena doble marcado radiactivamente (ADNcd) y el quinto el complejo THO:Tex1 con ADN de cadena simple marcado radiactivamente (ADNcs), el sexto el complejo THO obtenido a partir de la cepa THOΔTex1 (control negativo para el marcaje de autorradiografía), el séptimo el complejo THO incubado con ARN marcado radiactivamente, el octavo el complejo THO incubado con ADNcs marcado radiactivamente y el noveno el complejo THO incubado con ADNcd marcado radiactivamente.

A continuación se propuso identificar qué parte de la proteína Tho2 estaba involucrada en la interacción con los ácidos nucleicos. Para ello se diseñó en primer lugar la cepa mutante de *S. cerevisiae* Tho2-3C, en la que se introdujo un sitio de corte para la proteasa 3C entre los aminoácidos 1270 y 1271 con la intención de cortar el extremo carboxi-terminal de esta proteína (apartado 3.1.3). A continuación se purificó el complejo THO (apartado 3.1.2) y se incubó con la proteasa 3C y con ARN marcado radiactivamente, en condiciones idénticas a las descritas anteriormente. Durante el experimento se utilizó una muestra de complejo THO purificado a partir de la cepa original Tho2-TAP como control. La muestra se sometió a una

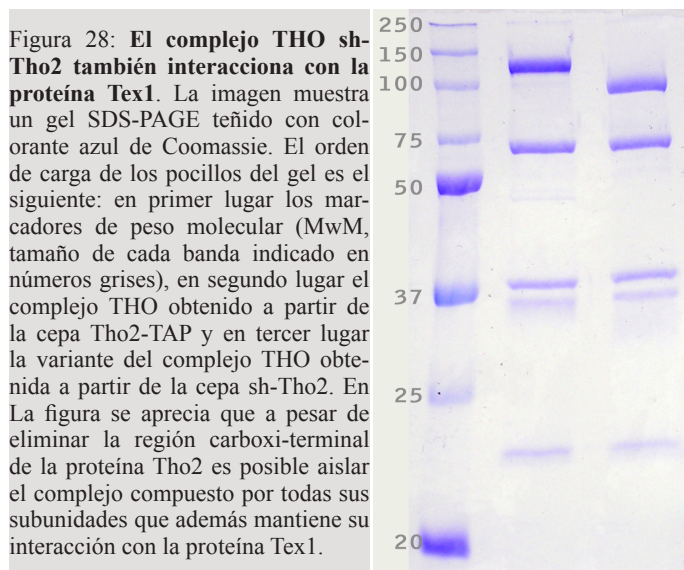
exposición de luz ultravioleta y se sometió a una electroforesis desnaturalizante en un gel SDS-PAGE (apartado 3.1.4). Después se introdujo el gel en un cámara de autorradiografía para detectar el ARN marcado radiativamente y se tiñó con colorante azul de Coomassie para identificar las proteínas presentes en el gel. El resultado del experimento mostró que sólo el complejo THO original y no el sometido al corte de la proteasa 3C en el extremo carboxi-terminal de la proteína Tho2 es capaz de interaccionar con el ARN. Este resultado sugiere que el complejo THO requiere el extremo carboxi-terminal de la proteína Tho2 para interaccionar con el ARN marcado radiativamente (Figura 27).

Además, los resultados también sugieren que el sito de corte de la proteasa 3C (en el extremo carboxi-terminal de la proteína Tho2) debe quedar expuesto al medio en la forma nativa del complejo, dado que la incubación con la proteasa 3C



se realizó después de purificar el complejo y la proteasa fue capaz de realizar el corte de la proteína, como revela la disminución del tamaño de Tho2 después del incubación con la proteasa 3C (Figura 27).

En base a estos resultados, se decidió generar una nueva cepa mutante (sh-Tho2) en la que se eliminó el dominio carboxi-terminal de la proteína Tho2 a partir del punto de corte de la proteasa 3C de la cepa Tho-3C (apartado 3.1.3). La cepa sh-Tho2 mostró una notable disminución en la tasa de crecimiento y los cultivos necesitaron mayor cantidad de tiempo para crecer, alcanzando una densidad óptica no superior a 1,2. El complejo se purificó siguiendo el protocolo habitual de purificación en tándem por afinidad (apartado 3.1.2) pero el rendimiento de la purificación fue bastante inferior al obtenido con de las cepas Tho2-TAP, THO Δ Tex1 y Tho2-3C. El resultado de la purificación mostró finamente un complejo íntegro y aparentemente normal a juzgar por su comportamiento en la cromatografía (apartado 3.1.2) y la observación por ME (apartado 3.6). Además, el análisis en electroforesis desnaturante mostró que la proteína Tho2 de este nuevo complejo THO (THO sh-tho2) presentaba un tamaño menor que la original. Estos resultados demuestran que es posible purificar el complejo THO en ausencia del dominio carboxi-terminal de la proteína Tho2. Además, se compró que la proteína Tex1 también copurifica con este complejo THO truncado en la proteína Tho2, lo que sugiere que esta región no interviene en la interacción entre el complejo THO y la proteína Tex1 (Figura 28).



Con el objetivo de comprobar que el complejo THO truncado en el extremo carboxi-terminal de la proteína Tho2 era incapaz de interactuar con los ácidos nucleicos, se repitieron los ensayos realizados anteriormente incubando la muestra procedente de la purificación a partir de la cepa sh-Tho2 con ADNcs, ADNcd y ARN marcados radiactivamente, manteniendo las mismas condiciones que en los experimentos anteriores. Los resultados mostraron que el complejo THO truncado en el extremo carboxi-terminal (THO sh-Tho2) no es capaz de interactuar con ninguno de los ácidos nucleicos testados, lo que sugiere que es posible generar y purificar una variante del complejo THO estable e incapaz de interactuar con los ácidos nucleicos. Al mismo tiempo estos resultados corroboran los obtenidos en el experimento anterior (con la cepa Tho2-3C) y evidencian el hecho de que el dominio carboxi-terminal de la proteína Tho2 es imprescindible para la interacción del complejo THO con los ácidos nucleicos ADNcs, ADNcd y ARN (Figura 29).

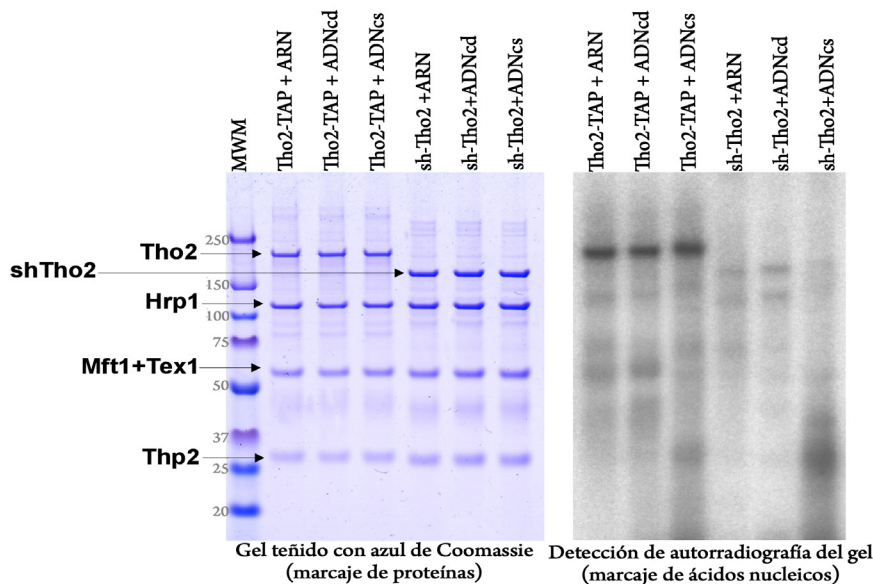


Figura 29: El complejo THO sh-Tho2 no interacta con los ácidos nucleicos. Resultado de someter el mismo gel (SDS-PAGE) a tinción con colorante azul de Coomassie para identificar las proteínas (izquierda) y a detección de autorradiografía para detectar los ácidos nucleicos marcados radiactivamente (derecha). El orden de carga de los pocillos del gel es el siguiente: en primer lugar los marcadores de peso molecular (MwM, tamaño de cada banda indicado en números grises), en segundo lugar el complejo THO obtenido a partir de la cepa Tho2-TAP incubado con ARN marcado radiactivamente, en tercer lugar el mismo complejo incubado con ADNcd marcado radiactivamente, en cuarto lugar el mismo complejo incubado con ADNcs marcado radiactivamente, el quinto lugar la nueva variante del complejo THO obtenida a partir de la cepa sh-Tho2 incubado con ARN marcado radiactivamente, en sexto lugar la nueva variante del complejo THO incubada con ADNcd marcado radiactivamente y en séptimo lugar la nueva variante del complejo THO incubada con ADNcs marcado radiactivamente. Se aprecia que la proteína Tho2, es mas pequeña en las muestras obtenidas a partir de la cepa sh-Tho2 que en las procedentes de la cepa Tho2-TAP y que sólo el complejo THO purificado a partir de la cepa Tho-TAP y no el procedente de la cepa sh-Tho2 interacta con los ácidos nucleicos.

4.11. Estudio fenotípico de las cepas mutantes en el complejo THO

Durante esta tesis doctoral se han utilizado complejos THO procedentes de distintas cepas mutantes de *S. cerevisiae* (apartado 3.1.3). A continuación se propuso identificar algunos de los fenotipos de estas cepas y para ello estudió el comportamiento de las células en su crecimiento y en su capacidad de expresar el gen lac-Z a diferentes temperaturas.

Para analizar las características fenotípicas de las cepas mutantes en el complejo THO se diseñó una nueva cepa para utilizarla como control negativo de la expresión del complejo THO. En base a estudios previos que demuestran que la eliminación de cualquier componente del complejo THO anula la formación del complejo, se diseñó la cepa mutante THO Δ Tho2, en la que se eliminó por completo el gen de la proteína Tho2 con el objetivo de evitar la formación del complejo THO (Huertas *et al.*, 2006). El hecho de que esta cepa fuese viable se debe a que el complejo THO no es esencial para la supervivencia celular de *S. cerevisiae* en condiciones óptimas (Jimeno *et al.*, 2002). Como control positivo para los ensayos fenotípicos se utilizó la cepa original Tho-TAP que contiene el complejo THO totalmente activo. Así pues, el estudio fenotípico se realizó atendiendo al comportamiento de cuatro cepas: Tho-TAP (control positivo, presenta un complejo THO totalmente activo), Tho2 Δ Tho2 (control negativo, el complejo THO no se forma), sh-Tho2 (el complejo THO carece del dominio carboxi-terminal de la proteína Tho2 y es incapaz de interaccionar con los ácidos nucleicos) y THO Δ Tex1 (presenta un complejo THO activo pero carecen de la proteína Tex1).

El crecimiento celular de cada cepa se observó a 30 °C (temperatura óptima de crecimiento) y a 37 °C (temperatura restrictiva). Para ello se prepararon varias diluciones de cada cepa, de forma que cada dilución contenía diez veces mayor cantidad de células que la anterior (apartado 3.4). El resultado mostró que, sorprendentemente, las células cuyo complejo THO carece del extremo carboxi-terminal de la proteína Tho2 (cepa sh-Tho2) mostraban prácticamente las mismas dificultades para crecer a la temperatura restrictiva (37 °C) como aquellas que carecen totalmente del complejo THO (cepa THO Δ Tho2). Por el contrario, la cepa THO Δ Tex1 no mostró

ninguna diferencia con respecto a la cepa Tho-TAP en el crecimiento a ninguna de las temperaturas testadas. Estos datos sugieren que el extremo carboxi-terminal de la proteína Tho2 desempeña un papel importante para la supervivencia celular de *S. cerevisiae* a 37 °C (Figura 30) mientras que la proteína Tex1 influye en el crecimiento ni en la supervivencia celular de *S. cerevisiae*. Este resultado es consistente con los datos previos que proponen que la ausencia de Tex1 no interviene en ninguna de las fases de la expresión génica (Luna *et al.*, 2005; García-Rubio *et al.*, 2008). Además, en este experimento se observó que la cepa THO Δ Tho2 es incapaz de crecer a 37 °C, lo que sugiere que el complejo THO se hace imprescindible para la supervivencia celular a 37 °C. Este dato también concuerda con los resultados previos expuestos en la bibliografía que afirman que el complejo THO es esencial para la supervivencia celular de *S. cerevisiae* y de *D. melanogaster* en condiciones de estrés térmico (Chávez *et al.*, 2000; Jensen *et al.*, 2001a; Strasser *et al.*, 2002; Rehwinkel *et al.*, 2004) (Figura 30).

La capacidad de expresión del gen lac-Z de las diferentes cepas se midió mediante el registro del producto de expresión del gen, la β -galactosidasa. El ensayo se realizó midiendo dicha actividad en cultivos a 30 °C y a 37 °C. Los resultados mostraron que la expresión del gen quedaba reducida prácticamente a la mitad en las células de la cepa sh-Tho2 a 30 °C mientras que el mismo efecto se hacía aún mas evidente a 37 °C, donde la expresión del gen en estas células representaba aproximadamente una cuarta parte de la expresión registrada en las células de la cepa Tho-TAP (Figura 30). Estos datos sugieren que el extremo carboxi-terminal de la proteína Tho2 debe representar un papel importante para la expresión de este gen, siendo su función especialmente importante a 37 °C. En el caso de las células THO Δ Tex1 se observó que la ausencia de la proteína Tex1 no afecta a expresión del gen lac-Z a 30 °C pero si a 37 °C, donde quedaba reducida aproximadamente a la mitad. Estos resultados sugieren que la proteína Tex1 no es determinante para la expresión del gen lac-Z a 30 °C pero si facilita su expresión a 37 °C. Además, el experimento mostró que, de acuerdo con estudios previos, las células que carecen del complejo THO son absolutamente incapaces de expresar el gen lac-Z a ninguna de las temperaturas contempladas (West *et al.*, 2000; Chávez *et al.*, 2001) (Figura 30).

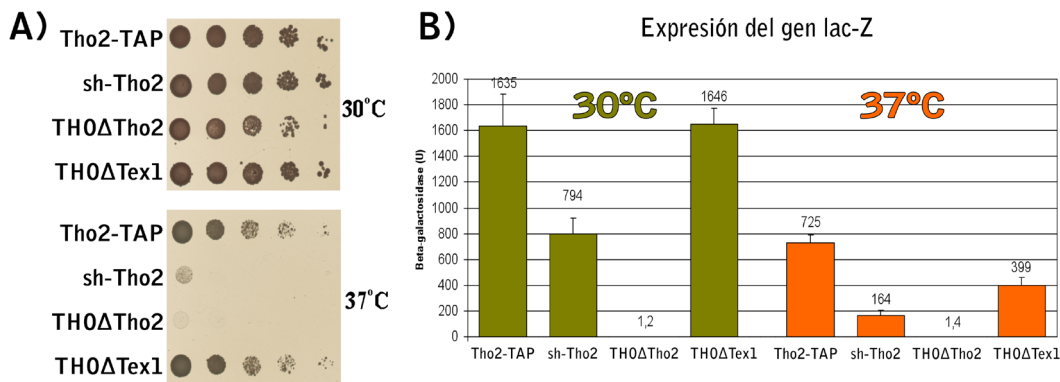
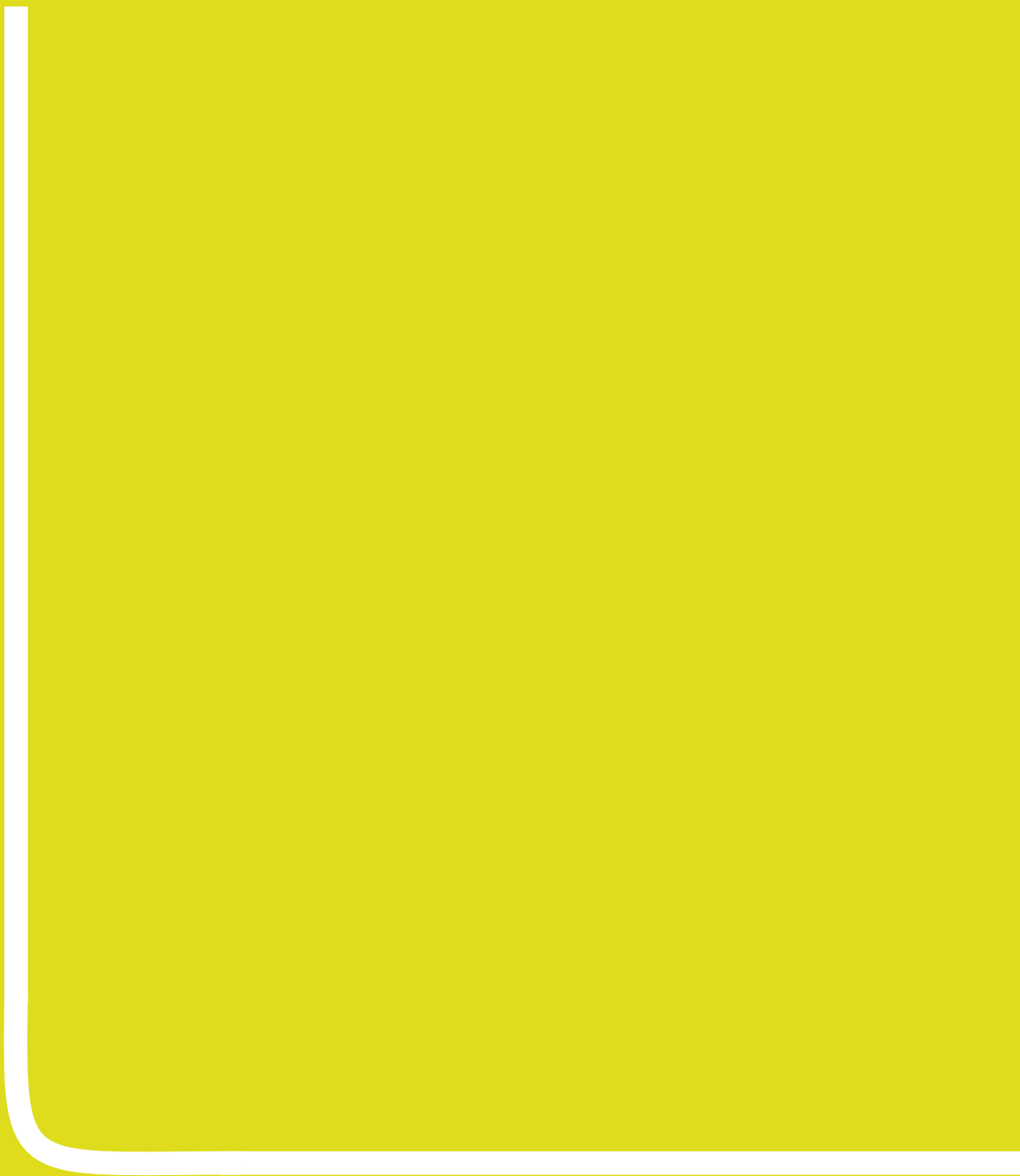


Figura 30: **Análisis del crecimiento y de la expresión del gen lac-Z en las diferentes cepas de *S. cerevisiae*.** **A)** La figura muestra el resultado de la observación del crecimiento de las cepas Tho2-TAP, sh-Tho2, THOΔTho2 y THOΔtex1 a 30 °C (arriba) y a 37 °C (abajo). Se observa que, mientras que a 30°C todas las cepas presentan un crecimiento aproximadamente similar, a 37 °C la cepa Tho2-TAP crece con mayor dificultad, la sh-Tho2 es prácticamente incapaz de crecer, la cepa THOΔTho2 no crece y la cepa THOΔTex1 crece de forma similar a la cepa Tho2-TAP. **B)** La imagen muestra un diagrama de barras obtenido a partir del registro en la expresión del gen lac-Z (producción de β-galactosidasa) en las diferentes cepas mutantes. Se observa que a la temperatura restrictiva (37 °C) la expresión del gen lac-Z se ve disminuida en todas las cepas excepto en la cepa THOΔTho2 que es incapaz de producir β-galactosidasa a ambas temperaturas. La cepa Tho2-TAP reduce la expresión del gen lac-Z aproximadamente a la mitad al pasar de 30 a 37 °C. La cepa sh-Tho2 expresa con mayor dificultad que la cepa Tho2-TAP el gen lac-Z a 30 °C y el efecto se hace aún mas evidente a 37 °C. La cepa THOΔTex1 muestra la misma capacidad de expresión del gen lac-Z a 30 °C que la cepa Tho2-TAP mientras que a 37 °C la expresión del gen se ve fuertemente reducida.



5. Discusión



5.1. El concepto “TREX”

En el año 2002 se utilizó por primera vez el concepto TREX para describir a un complejo muy conservado en la evolución desde levaduras hasta humanos implicado en la transcripción y exportación del ARNm. El complejo TREX se definió como la interacción entre el complejo THO, la helicasa Sub2 (UAP56 en humanos), la proteína de unión al ARNm Yra1 (Aly en humanos) y la proteína Tex1, hasta entonces desconocida (Strasser *et al.*, 2002) (Figura 31).

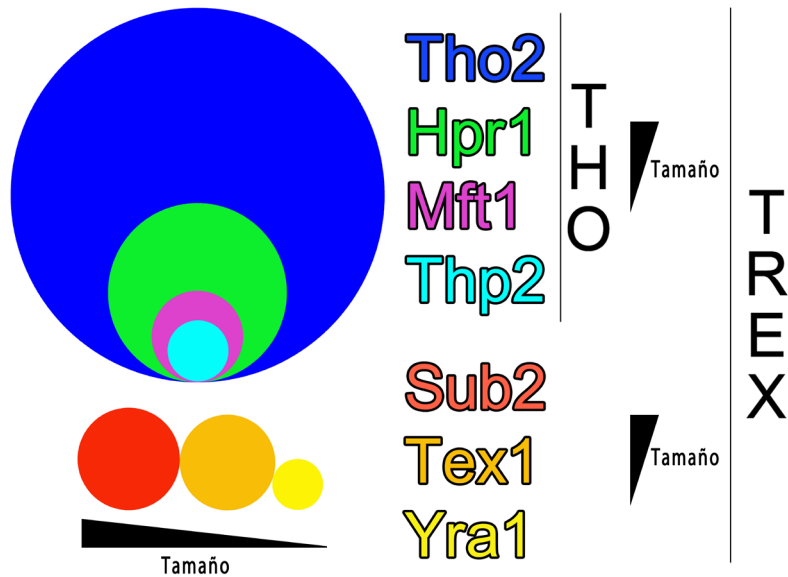


Figura 31: **Componentes del complejo TREX de *S. cerevisiae*.** El complejo TREX de *S. cerevisiae* se define como un hetero-heptámero compuesto por las cuatro subunidades del complejo THO (Tho2, Hpr1, Mft1 y Thp2) y las proteínas Sub2, Tex1 e Yra1. La figura muestra las siete subunidades del complejo TREX agrupadas en dos líneas, de forma que el complejo THO aparece en la parte superior y los tres nuevos componentes del complejo TREX en la parte inferior. El tamaño de cada proteína es, de mayor a menor es: Tho2 183 kDa, Hpr1 87,8 kDa, Sub2 50,3 kDa, Tex1 47,2 kDa, Mft1 44,9 kDa, Thp2 30,3 kDa e Yra1 24,9 kDa. La proteína Tho2 es también la mas grande del complejo TREX y representa aproximadamente un 40% del tamaño total del complejo. En la figura cada proteína está representada por una circunferencia cuyo área es proporcional al tamaño real de la proteína.

Así, apenas dos años después de su descubrimiento, el complejo THO se presentó como parte de un complejo de mayor tamaño en el que desempeña un papel protagonista. En el complejo TREX, THO se relacionó con el reclutamiento y carga de los factores de exportación Sub2 e Yra1 sobre el ARNm durante la elongación transcripcional (Reed *et al.*, 2005). De las tres proteínas que interaccionan con el complejo THO para formar el complejo TREX, Sub2 parece ser la mas im-

DISCUSIÓN

portante (García-Rubio *et al.*, 2008). Se trata de una proteína esencial, con un peso molecular de aproximadamente 50 kDa que interacciona directamente con Hpr1 en el complejo THO, con Yra1 del complejo TREX y con el exosoma nuclear, con el que parece colaborar con los mecanismos de control de calidad del ARNm (Jensen *et al.*, 2001a; Libri *et al.*, 2002). Sub2 presenta una función bastante similar a la descrita para el complejo THO y de hecho, se ha demostrado que la sobreexpresión de Sub2 suprime el fenotipo de las células deficientes en el complejo THO, lo que sugiere que en determinadas condiciones, el exceso de esta proteína puede suplir la función del complejo THO (Fan *et al.*, 2001). Por el contrario, Yra1 es un factor de exportación pequeño (25 kDa) que interacciona directamente con el ARNm (Portman *et al.*, 1997), con el factor de exportación Mex67 (Strasser *et al.*, 2000a; Zenklusen *et al.*, 2001) y con Sub2, que funciona como adaptadora de esta proteína al complejo TREX (Zenklusen *et al.*, 2001; Gwizdek *et al.*, 2006; Hautbergue *et al.*, 2008). Además, Yra1 también interacciona directamente con el exosoma por lo que también ha sido relacionada con la maquinaria de control de calidad de ARNm (Zenklusen *et al.*, 2002).

El complejo TREX no presenta ninguna función diferente a las descritas para complejo THO. Dado que las interacciones entre los componentes del complejo THO son mucho más fuertes que aquellas que se producen entre los componentes del complejo TREX (Strasser *et al.*, 2002; Hurt *et al.*, 2004; Reed *et al.*, 2005; Huerfías *et al.*, 2006) se ha propuesto que el complejo THO recluta Sub2 e Yra1 durante la transcripción y los carga directamente sobre el ARNm, contribuyendo así a la biogénesis de RNPm (Reed *et al.*, 2005). Según esta definición, se podría considerar que el llamado complejo TREX no es sino la forma activa del complejo THO.

A medida que avanza la comprensión y el estudio del proceso de biogénesis de RNPm se han ido identificando nuevas proteínas que interaccionan directamente con el complejo THO (Reed *et al.*, 2005). Durante la elongación transcripcional el complejo THO recorre el gen unido a la ARNP_{II} y su función parece estar directamente relacionada con su capacidad de interaccionar simultáneamente con varios factores de exportación, la mayoría de los cuales son posteriormente cargados sobre el ARNm recién sintetizado (Reed *et al.*, 2005). Los factores de exportación Gbp2

y Hrb1 son componentes de los RNPm que son incorporados sobre el ARNm recién sintetizado por medio de su interacción con las proteínas Hpr1 y Mft1 del complejo THO (Hacker *et al.*, 2004; Luna *et al.*, 2005). Tanto Gbp2 como Hrb1 copurifican con el resto de subunidades del complejo TREX (Hurt *et al.*, 2004). El factor de exportación Mex67 también interacciona con la proteína Hpr1 del complejo THO y es posteriormente cargado sobre el ARNm recién sintetizado. Mex67 protege al complejo THO de la degradación y se le considera responsable de la estabilidad general del complejo THO *in vivo* (Kadowaki *et al.*, 1994; Gwizdek *et al.*, 2006).

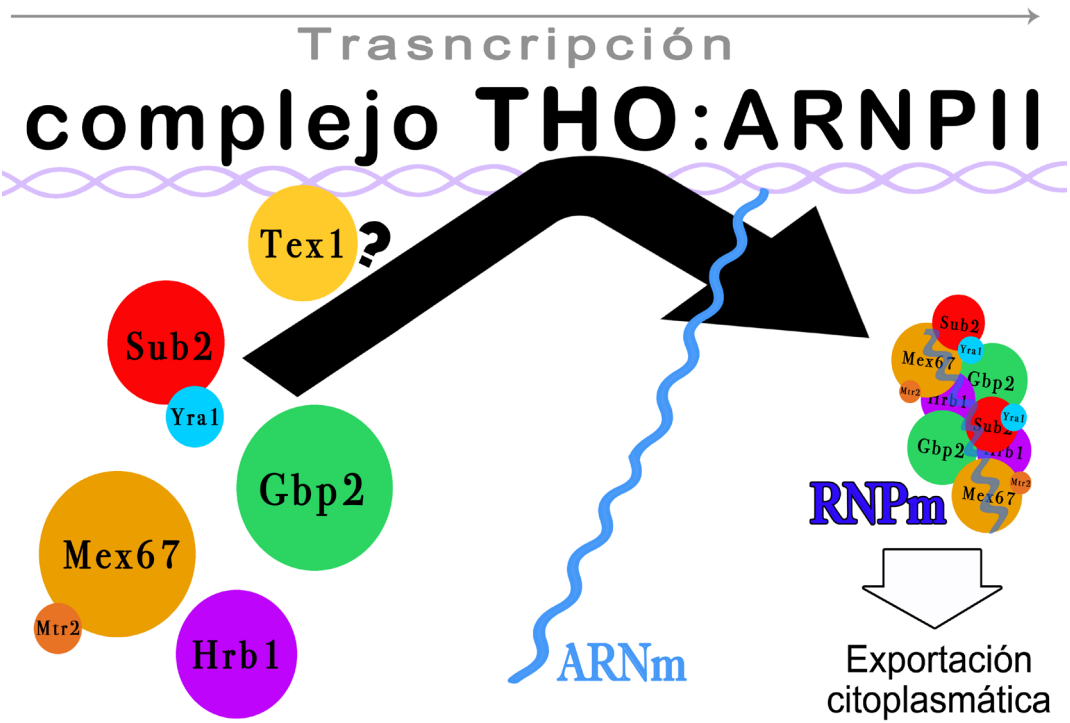


Figura 32: **Proteínas que interaccionan con el complejo THO de *S. cerevisiae*** durante la realización de su función fisiológica. Durante la transcripción el complejo THO permanece asociado a la ARNPm formando el complejo THO:ARNPII, que recorre el gen al tiempo que sintetiza la molécula de ARNm (ADN representado en morado y ARNm en azul). La mayoría de las proteínas que interaccionan directamente con el complejo THO son cargadas en el ARNm (interacción representada por la flecha negra) y pasan a formar parte de los RNPm, que son posteriormente exportados al citoplasma celular. El complejo TREX fue definido por la interacción entre las proteínas Sub2, Yra1, Tex1 y el complejo THO. En la figura se muestra que la función general de los factores Sub2, Yra1, Mex67, Mtr2, Gbp2 y Hrb1 desempeñan una función genreal similar, dado que todas son cargadas sobre el ARNm por acción del complejo THO, mientras que Tex1 permanece unida al complejo desempeñando una función desconocida (la interrogación jutno a Tex1 hace referencia a su función desconocida). Así, el concepto TREX define un complejo que mezcla proteínas que son cargadas en el ARNm con otras que no lo son y excluye otras que contribuyen de igual forma a la realización de la función fisiológica del complejo THO en la biogénesis y exportación citoplasmática de información genética.

Si tenemos en cuenta que el complejo TREX hace referencia al un complejo implicado en la “transcripción-exportación” de ARNm, que puede definirse como la forma activa del complejo THO y que tanto Gbp2, como Hrb1 y Mex67 interactúan con el complejo THO durante la transcripción contribuyendo a su función, quizá se podría considerar a estas proteínas como componentes del complejo TREX (Hurt *et al.*, 2004; Reed *et al.*, 2005; Saguez *et al.*, 2008). Otra posibilidad sería definir una serie de complejos que se forman con carácter temporal durante la elongación transcripcional, de forma que las proteínas interactúan en primer lugar con el complejo THO y a continuación quedan ancladas al ARNm. La interacción de Gbp1, Hrb1, Sub2, Yra1 y Mex67 con el complejo THO podrían definir diferentes complejos TREX, que se podrían llamar TREX1, TREX2, TREX3, etc. y que se forman con carácter temporal a lo largo del proceso de transcripción. Como alternativa, se podría quizá suprimir el término TREX o utilizarlo como sinónimo de THO, dado que la definición actual del concepto TREX responde básicamente a la forma activa del complejo THO (Figura 32).

Los resultados presentados en esta tesis doctoral demuestran que la proteína Tex1 interactúa de forma muy estable con el complejo THO. Sin embargo, en la definición actual del complejo TREX, Tex1 parece ser la proteína menos importante y con menor relevancia para la función del complejo. Dado que tanto Yra1 como Sub2 interactúan con el ARNm y forman parte del RNPm, la función del complejo TREX se atribuye al reclutamiento de factores sobre el ARNm, pero la proteína que interactúa más estrechamente con el complejo THO no parece tener ninguna implicación directa con el ARNm (Rehwinkel *et al.*, 2004). El hecho de que Tex1 sea la proteína mas estrechamente unida al complejo THO ha provocado que algunos autores hayan considerado al complejo THO:Tex1 como el núcleo del complejo TREX (Oppizzi, 2004). La caracterización de la función de la proteína Tex1 será decisiva para la determinación de la función del complejo TREX.

5.2. El complejo THO y las proteínas WD

La purificación realizada a partir de la cepa Tho2-TAP ha dado como resultado la obtención del complejo THO asociado a Tex1, una proteína de la familia de las WD (apartado 4.2). Esta interacción ha sido previamente descrita por varios autores (Jimeno *et al.*, 2002; Strasser *et al.*, 2002; Rehwinkel *et al.*, 2004; García-Rubio *et al.*, 2008) e incluso, algunos experimentos han mostrado que Tex1 interacciona con el complejo THO de forma independiente de Sub2, de Yra1 y de los ácidos nucleicos (Strasser *et al.*, 2002; García-Rubio *et al.*, 2008). La interacción entre THO y Tex1 ha sido descrita en *S. cerevisiae* (Strasser *et al.*, 2002), *D. melanogaster* (Rehwinkel *et al.*, 2004) y en *H. sapiens* (Strasser *et al.*, 2002). Este hecho demuestra que la interacción THO:Tex1 esta muy conservada en la evolución, lo que sugiere que la función de Tex1 junto al complejo THO desempeña un papel importante para la supervivencia de los organismos (Reed *et al.*, 2005).

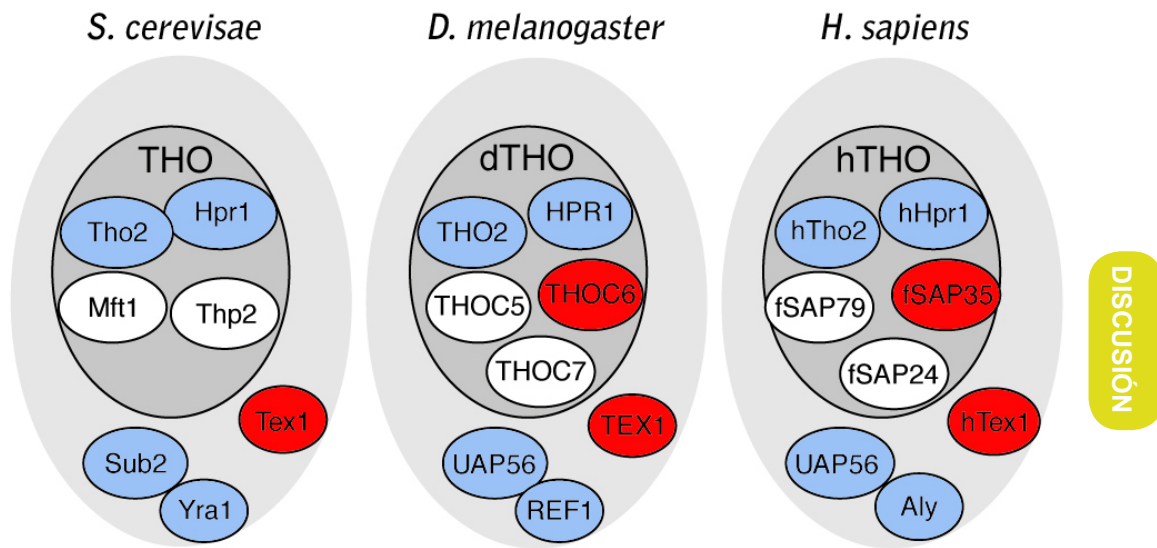


Figura 33: **Componentes de los complejo THO y TREX conservados en la evolución.** La figura representa los componentes de los complejos THO y TREX tal y como han sido definidos en *S. cerevisiae*, *D. melanogaster* y *H. sapiens*. Se han coloreado en rojo las proteínas que presentan una estructura en β -propeller y pertenecen a la familia de las proteínas WD40. Además, en azul se han coloreado las proteínas correspondientes a los complejos THO y TREX que presentan homólogos en las demás especies y por tanto, están conservadas a lo largo de la evolución. La proteína Tex1 no está coloreada en azul pero también está conservada a lo largo de la evolución. Los factores de exportación Mex67 y Gbp2 no aparecen en la figura porque no han sido descritos como componentes del complejo TREX. Figura adaptada de Reed *et al.*, 2005.

Además de la proteína homóloga de Tex1 (Thoc3 ó hTex1) en el complejo THO de *H. sapiens* y *D. melanogaster*, existe otra proteína llamada Thoc6 (*THO complex subunit 6*) o fSAP35 (*functional spliceosome-associated protein 35*) que no tiene homólogo en levaduras y que también pertenece a la familia de las proteínas WD. Así, mientras que en levaduras sólo se ha descrito una proteína de la familia WD implicada en la función del complejo THO, en *D. melanogaster* y *H. sapiens* se han descrito dos, la homóloga de Tex1 y Thoc6 (Figura 33).

La interacción entre la proteína Tex1 y el complejo THO es significativamente mas estable que con el resto de proteínas asociadas al complejo, tanto en *S. cerevisiae* como en *D. melanogaster* y en *H. sapiens* (Strasser *et al.*, 2002; Rehwinkel *et al.*, 2004). Asimismo, en el complejo THO de *H. sapiens* y *D. melanogaster* la proteína WD Thoc6 no aparece asociada al complejo sino como componente intrínseco de éste. Sin embargo, estos datos contrastan con los resultados presentados en esta tesis doctoral y en estudios previos que muestran que la eliminación de la proteína Tex1 del genoma de *S. cerevisiae* no provoca cambios aparentes en el crecimiento de las células de *S. cerevisiae* (Figura 30) y apenas afecta a la expresión del gen *laz-Z* incluso a temperatura restrictiva (Luna *et al.*, 2008).

Llegado este punto surge la duda acerca de la función que desempeñan las proteínas WD en la función del complejo THO. Tex1 es el único componente asociado al complejo THO que no parece interaccionar con el ARNm ni con el ADN (Zenklusen *et al.*, 2002; Hurt *et al.*, 2004). La predicción de estructura tridimensional de la proteína Tex1 revela la presencia de repeticiones de tipo WD40 organizadas en una estructura tridimensional con plegamiento en β -propeller en la región amino-terminal (Figura 23). La familia de proteínas WD se extiende a lo largo de toda la evolución en células eucariotas y algunas de ellas han sido descritas funcionando en la regulación del proceso de transcripción (Thomas *et al.*, 1995; De Silva *et al.*, 1998; Pickles *et al.*, 2002). La función que comparten las proteínas WD está relacionada con la facilitación de las interacciones proteína-proteína y la coordinación del ensamblaje de complejos multiproteicos, donde las repeticiones WD ejercen de esqueleto rígido para la interacción de las mismas (Smith *et al.*, 1999; Li *et al.*, 2001). Curiosamente, pese a no ser una proteína de la familia WD, Tho2, la proteína

mas grande del complejo THO, presenta un tipo de repeticiones en su secuencia de aminoácidos conocidas como motivos KEKE que también se han relacionadas con el ensamblaje de complejos y con la asistencia de interacciones proteína-proteína (Wall, 1973; Realini *et al.*, 1994; West *et al.*, 2000). El hecho de que la proteína homóloga de Tho2 en eucariotas superiores también presente este tipo de repeticiones y que el complejo THO de *D. melanogaster* y *H.sapiens* esté compuesto además por una proteína WD (Thoc6) y asociado de forma estable con otra proteína WD (Thoc3) es consistente con la función del complejo THO como mediador del proceso de biogénesis de RNPM responsable de la interacción con múltiples factores de exportación a lo largo de su función (Reed, 2003).

En el análisis estructural de los complejos THO y THO:Tex1 se ha visto que la proteína Tex1 interacciona con el complejo THO en una posición bastante céntrica con respecto al eje mayor del complejo, aproximadamente a la mitad de distancia entre ambos extremos. Esta localización podría suponer una posición estratégica para favorecer las interacciones del complejo THO con otros factores de exportación e incluso para facilitar en ensamblaje del propio complejo. Además, Tex1 interacciona con el complejo THO en una posición ligeramente desplazada hacia la proteína Hpr1, que ha sido propuesta como la responsable de la formación del

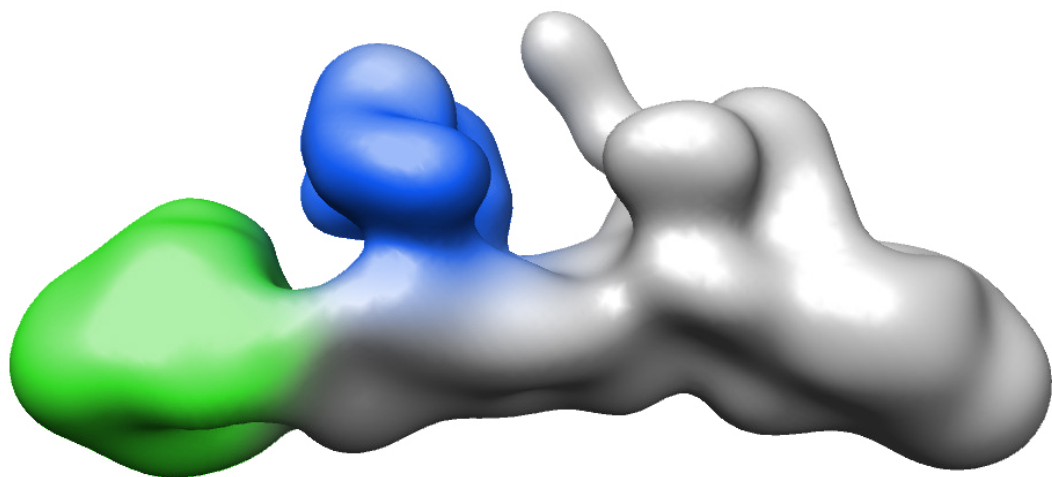


Figura 34: **Localización de las proteínas Hpr1 y Tex1 en el complejo THO:Tex1.** La figura muestra la localización de las proteínas Hpr1 (en verde) y Tex1 (en azul) en la reconstrucción tridimensional del complejo THO:Tex1. Se aprecia que la proteína WD Tex1 queda localizada en el complejo cerca de la proteína Hpr1, que ha sido identificada como responsable de la interacción con la ARNPII (Chang *et al.*, 1999) y con prácticamente todos los factores de exportación con los que interacciona el complejo THO (Zenklusen *et al.*, 2002; Hacker *et al.*, 2004).

complejo THO (Gwizdek *et al.*, 2005). Hpr1 es además el componente del complejo THO que ha sido identificado como responsable de la interacción con mayor número de proteínas (interacciona con la ARNP_{II}, Sub2, Mex67, Hrb1 y Gbp2) (Chang *et al.*, 1999; Chávez *et al.*, 2000; Zenklusen *et al.*, 2002; Hacker *et al.*, 2004; Gwizdek *et al.*, 2006). Así, sería lógico pensar que la localización de la proteína Tex1 en el complejo guarde relación con su función en el mismo y visto que aparece localizada cerca de Hpr1, se podría pensar que colabora en la oligomerización del complejo THO, en la interacción con la ARNP_{II} y en el reclutamiento cotranscripcional de factores de exportación (Figura 34).

El modelo atómico de la proteína Tex1 presenta un plegamiento en β -*propeller* en la región amino-terminal. La estructura en β -*propeller* representa aproximadamente el 78% de la secuencia completa de la proteína, lo que sugiere que Tex1 tiene además una región carboxi-terminal (22% de la proteína) cuya estructura tridimensional no ha sido incluida en el modelo atómico. La predicción de estructura secundaria muestra que la región carboxi-terminal no incluida en el β -*propeller* responde a una estructura rígida con plegamiento de tipo lámina β . Esta observación es consistente con los resultados observados en el ajuste de la modelo de la región amino-terminal de Tex1 en la estructura tridimensional del complejo THO:Tex1, en la que se aprecia una masa estable contigua al modelo en β -*propeller* (Figura 35).

Así, la proteína Tex1 podría estar compuesta por dos dominios, un amino-terminal que adopta estructura en β -*propeller* con ocho hojas β y otro carboxi-terminal con una estructura mas o menos rígida (láminas β). Dado que la especificidad de las proteínas de la familia WD viene generalmente determinada por la parte de la proteína que queda fuera de las repeticiones WD (Neer *et al.*, 1994), la caracterización de este dominio podría ser decisiva para la caracterización funcional de esta proteína en el proceso de transcripción y en la exportación de RNPm.

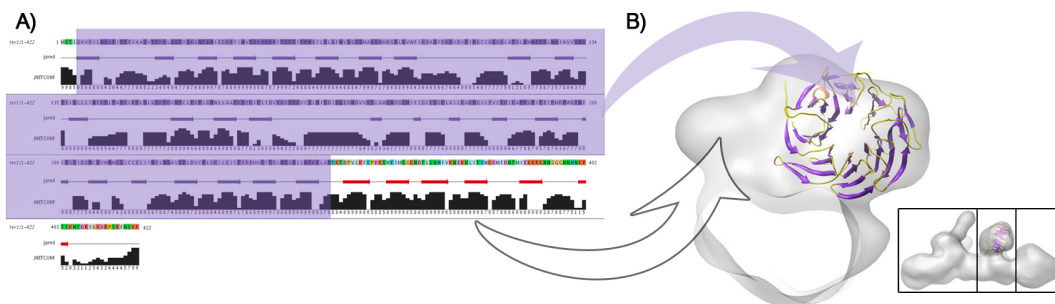


Figura 35: Predicción de la estructura secundaria de Tex1 en las regiones amino y carboxi-terminal. A) El modelo atómico generado computacionalmente para la proteína Tex1 comprende desde el aminoácido 5 hasta el 337 de los 422 totales que componen la proteína completa y la predicción obtenida sugiere que esta parte de la secuencia adopta un plegamiento en β -propeller. En la figura se han marcado sobre fondo morado los aminoácidos que se encuentran incluidos en el β -propeller. Sin embargo, a la estructura en β -propeller la suceden otros ochenta aminoácidos que también parecen responder a una estructura secundaria rígida, organizada en láminas β (flechas rojas en el parámetro jpred). Estos aminoácidos podrían estar formando un dominio carboxi-terminal en la proteína Tex1. B) La figura muestra la sección del volumen del complejo THO:Tex1 correspondiente a la proteína Tex1 con el modelo atómico del dominio amino-terminal de Tex1 ajustado en su interior. Se observa que el volumen del complejo THO:Tex1 identificado como correspondiente a la proteína Tex1 es mas grande que el modelo atómico de la región amino-terminal de la proteína, lo que genera un espacio vacío en el volumen junto a la estructura en β -propeller. Este espacio podría corresponder al posible dominio carboxi-terminal de Tex1, cuya estructura no está incluida en el β -propeller. Las flechas representan la posible distribución de la secuencia de Tex1 en el volumen del complejo THO:Tex1 correspondiente. A la derecha se ha representado la sección del complejo THO:Tex1 en la que se ha ajustado el modelo de Tex1, delimitado por líneas negras.

5.3. Tho2 en la interacción con los ácidos nucleicos

La observación de que el extremo carboxi-terminal de la proteína Tho2 es responsable de la interacción con los ácidos nucleicos es consistente con la alta cantidad de Argininas presentes en esta región de su secuencia. Las Argininas son aminoácidos fuertemente básicos que se encuentran presentes en muchas de las proteínas involucradas en todos los procesos que intervienen en la expresión génica (Hacker *et al.*, 2004; Reed *et al.*, 2005; Zhong *et al.*, 2009). Estudios previos han identificado dominios de interacción con el ARNm que utilizan sus regiones ricas en Argininas para interaccionar con los nucleótidos, demostrando que la abundancia de Argininas en una región determinada es suficiente para permitir la interacción de las proteínas con los ácidos nucleicos (Hautbergue *et al.*, 2008). Así, la interacción entre el extremo carboxi-terminal de la proteína Tho2 y los ácidos nucleicos estaría favorecida por la alta cantidad de Argininas presentes en esta región, que representan mas del 10% de los aminoácidos de su secuencia (Figura 36).

MAEQTLTLLSKLNLASQKVIIPASPSQASILTEEVIRNWPERSKTLCSDFTALESNDEKEDW
 LRTLFIELFDFINKNDENSPLKLSDVASFSTNELVNHERQVSQASIVGKMFIAVSSSTVPNI
 NDLTTTISLCKLIPSLHEELFKFSWISSKLLNKEQTTLRLHLLKSKYELKKYNLLVENS
 GYGQLVALLILAYDPDNFSKVSAYLKEIYHIMGKYSLDSIRTLDDVILNVSSQFITEGYK
 FFIALLRKSDSWPSSHVANNSSNYSSLNNEGGNMIAANISFNLSQYNEEVDKENYERYMDM
 CCILLKNGFVNFYSIWDNVKPEMEFLQEYIQNLETELEESTKGVENPLAMAAALSTENE
 TEDNALVVDNDVMKDKISEETNADIESKGGKQKQDILLFGKIKLLERLLIHGCVIPV
 IHVLKQYPKVLYVSELSRYLGRVFEYLLNPLYTSMTSSGESKDMATAMITRIDNGILA
 HKPRLIHKYKTHEPFESLELNSSYVFYSEWNSNLTPEASVNDLFENSHIYLSIIGPYLG
 RIPTLLSKISRIGVADIQKNHGESLHVTIDKWDIYVRKFIFPATSLLONNPIATSEVYE
 LMKFFPFKEKRYFIYNEMMTKLSQDILPLKVSFNKAEREAKSILKALSIDITAKESRRFAK
 LISTNPLASLVPAVKQIENYDKVSELVVYTTKYFNDFAVDVLQFVLLRLTYNRPVAVQFD
 GVNQAMWVQRLSIFITAGLAKNCPNMDISNIITYILKTLHNGNIIAVSILKELIITVGGIR
 DLNEVNMKQLMLNSGSPKQYARHLIYDFRDDNSVISRLTSFTDQSAISEIILLLYT
 LNLKANTQNSHYKILSTRCDEMNTLLWSFIELIKHCLKGKAFFENVLPFVELNNRPHLST
 PWTFHIWRDYLNDQLNSNENFSIDELIEGAEFSDVDLTKISKDLFTTFWRSLYDIHF
 SLYDERKNALSGENTGHMSNRKKHLIQNKIDILVTGISHQRAFKKTSEFISEKSNVWNK
 DCGEDQIKIFLQNCVVRVLFSPSDALFSSFFIFMAFRTENLMSILNTCITSNLIKTLTF
 CCTSSEAGNLGLFFTDVLKKLEKMRLNGDFNDQASRKLIEWHSVITEQVIDLLSEKNYMS
 IENGIEFMKHVTSVFPVVKAHIQLVYTTLEENLINEEREDIKLPSSALIGHLKARLKDAL
 ELDEFCTLTEEEAEQKRIREMELEEIKNYETACONEQKQVALRKQLELNKSQRLQNDPPK
 SVASGSAGLNSKDRYTSRNEPVIPTKPSSQWSYSKVTBHVDDINHYLATNHLQKAI
 VENDDETNRKLKSKQNMPIFDNRNSTLEIFERYFTLIQNPQNPDAEKIDSLKRYIKN
 ISSEPYPTDTSSYSEAAPEYTKRSSRYSGNAGGKDGYSNNYRGPSSNDSSAPKNIKPIS
 SYAHKRSELPTSPSKSTYNDRSRALEPTGPDRGDFDQDRNRLREEYKKNSSQSSQLRF
 PEKPFQEGKDSKANPYQASSYKRDSPSENEKPNKRFKKDETIRNKFTQDYRNTD
 AARANENQRYNGNRKSNTOALPQGPKGNGYVSRQY

Figura 36: Composición en Argininas de la proteína Tho2 de *S. cerevisiae*. La figura muestra la secuencia aminoacídica completa de la proteína Tho2 de *S. cerevisiae* en la que se han resaltado todas las Argininas. El dominio carboxi-terminal de Tho2, coloreado en verde, presenta las Argininas en color rojo furete, frente al color rojo mas claro del resto de Argininas de la secuencia. El recuento muestra que el 77% de las Argininas presentes en la secuencia se distribuyen a lo largo del dominio carboxi-terminal de la proteína, donde mas del 10% de los aminoácidos son Argininas.

Por otro lado, el hecho de que la eliminación del extremo carboxi-terminal de la proteína Tho2 provoque una disminución drástica en el crecimiento celular de *S. cerevisiae* a 37 °C (Figura 30) demuestra que esta región es casi tan importante como el resto del complejo THO para la supervivencia celular de *S. cerevisiae* en condiciones desfavorables. Sería razonable pensar que, al menos parte de la importancia de esta región, reside en su capacidad de interaccionar con los ácidos nucleicos. Dado que el complejo THO se ha visto directamente implicado en el mantenimiento de la estabilidad genética y en la abolición de la formación de híbridos ADN:ARN (Santos-Rosa *et al.*, 1994; Chávez *et al.*, 1997; Jimeno *et al.*, 2002; Huertas *et al.*, 2003) se podría pensar que la interacción del complejo con los ácidos nucleicos desempeña un papel protagonista en esta función. Así, en ausencia del complejo THO o de la región carboxi-terminal de la proteína Tho2 aumentaría la recombinación mitótica asociada a la transcripción y la inestabilidad genética,

repercutiendo probablemente en la supervivencia celular. Este modelo sugiere que la proteína Tho2 es la responsable de la estabilidad genética y es consistente con la observación de que las cepa mutantes en la proteína Hpr1 presentan defectos en la biogénesis de RNPM sin afectar a la estabilidad genética (Huertas *et al.*, 2006). Según este modelo, la proteína Hpr1 desempeñaría un papel primordial en la biogénesis de RNPM, interaccionando con múltiples factores de exportación, mientras que la proteína Tho2 sería responsable del mantenimiento de la estabilidad genética asociada a la transcripción, lo cual es consistente con la observación de que Tho2 y Hpr1 son los componentes mas importantes para la realización de la función fisiológica del complejo THO y que las cepas mutantes en la proteína Tho2 son mas sensibles a la exposición a luz ultravioleta que en las mutadas en la proteína Hpr1 (González-Barrera *et al.*, 2002). Esta propuesta sería también consistente con la reciente hipótesis que plantea en metazoos una posible diversificación de funciones de las diferentes subunidades del complejo THO en los distintos tejidos celulares (Jimeno *et al.*, 2010). En cualquier caso, la división de funciones entre los componentes del complejo no sería total, dado que el mutante sh-Tho2 también mostró problemas en la expresión génica, lo que sugiere que la región carboxi-terminal de Tho2 también interviene en otras funciones del complejo THO no relacionadas directamente con la estabilidad genética celular.

Sorprendentemente el extremo carboxi-terminal de la proteína Tho2 interacciona tanto con el ADN de cadena doble como con el ADN de cadena sencilla y el ARN. Teniendo en cuenta que durante la transcripción el complejo THO recorre los genes y evita la recombinación genética asociada a la transcripción y la formación de híbridos ADN:ARN (Santos-Rosa *et al.*, 1994; Piruat *et al.*, 1998; Schneiter *et al.*, 1999; Chávez *et al.*, 2000; Huertas *et al.*, 2003) podría quizá explicarse el hecho de que interaccione igualmente con ambos ácidos nucleicos. Así, el complejo THO podría interaccionar con cualquier gen y con cualquier ARN protegiéndolo y evitando las interacciones no deseadas entre cadenas nucleotídicas. La función de dicha interacción iría dirigida principalmente a la protección de las cadenas nucleotídicas y al mantenimiento de la estabilidad genética durante el proceso de transcripción (Luna *et al.*, 2005).

El hecho de que la purificación del complejo THO a partir de la cepa Tho2-TAP resultase exitoso es consistente con la posterior identificación de la región carboxi-terminal de la proteína Tho2 como región implicada de la interacción del complejo THO con los ácidos nucleicos ya que ambos datos apuntan a que esta región debe quedar expuesta al medio. Para localizar el extremo carboxi-terminal de la proteína Tho2, se reconstruyó el complejo resultante de la purificación a partir de la cepa sh-Tho2. A pesar de que el complejo THO carecía de aproximadamente el 10%

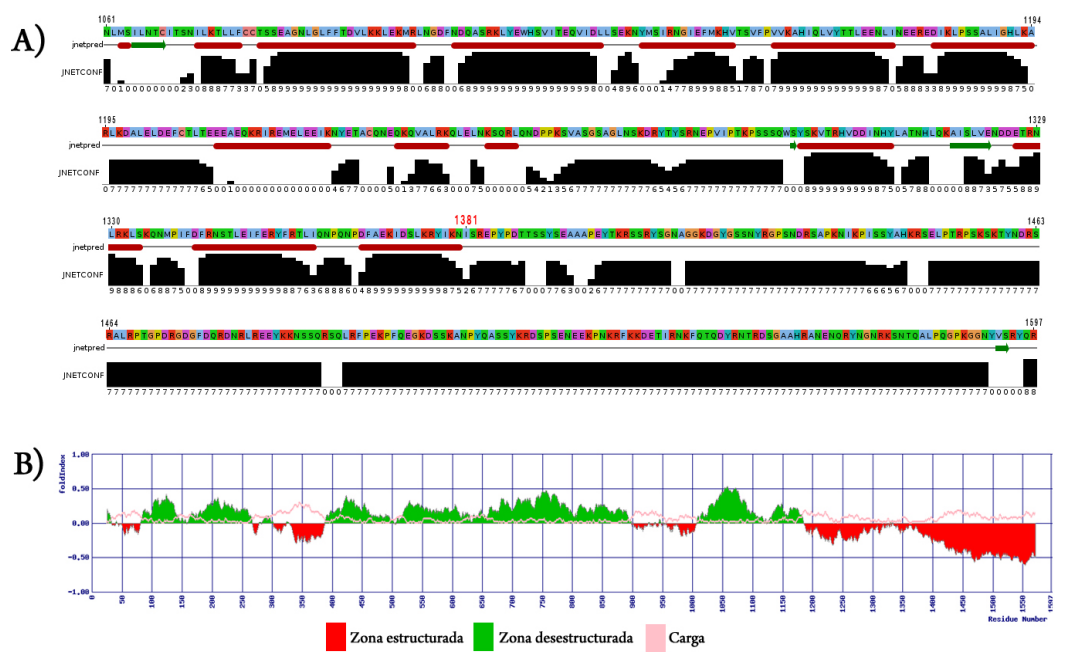


Figura 37: **El extremo carboxi-terminal de Tho2 está aparentemente desestructurado.** A) Predicción de estructura secundaria de la región carboxi-terminal de la proteína Tho2 realizada con el programa JPRED3 (Cole et al., 2008). La secuencia de Tho2 (1597 aminoácidos) aparece en la parte superior de cada fila, coloreada según el tipo de aminoácido. Los valores del parámetro “JPRED” corresponden al resultado final de la predicción, donde las zonas de la secuencia cuya predicción de estructura secundaria corresponde con el plegamiento en hélices α aparecen representadas mediante barras de color granate. Los valores del parámetro “JNETCONF” representan (en color negro) el valor de confianza de la predicción realizada, que oscila de 1 a 9 de forma que cuanto mayor es el valor mejor es la predicción. La figura muestra que casi toda la región carboxi-terminal, desde el aminoácido 1381 hasta el final de la proteína, se predice como región desestructurada. Los colores de los aminoácidos en la secuencia responden al siguiente esquema de colores: en azul claro Alanina, Isoleucina, Leucina, Metionina, Fenilalanina, Valina y Triptófano (los mas apolares); en verde Treonina, Glutamina, Asparagina y Serina (los mas hidrofílicos); en rosa el Ácido Glutámico y el Ácido Aspártico (cargados negativamente); en naranja Arginina y Lisina (cargados positivamente); en turquesa Tirosina e Histidina; en amarillo Prolina; en rojo claro Cisteina; en naranja claro Glicina. B) Predicción de plegamiento realizado con el programa FOLDINDEX (Prilusky et al., 2005) en el que se muestra en color verde las zonas de la secuencia con posible estructura secundaria y en rojo aquellas que se predicen como desestructuradas. La línea rosa muestra las zonas cargadas en la secuencia, atendiendo a los residuos de los aminoácidos que la componen. Se observa que toda la región carboxi-terminal, desde el aminoácido 1185 hasta el final, se predice como desestructurada. La región final de la secuencia presenta además una zona cargada positivamente, consistente con su capacidad de interactuar con los ácidos nucleicos.

de su masa total (la región carboxi-terminal de Tho2 tiene un tamaño de 35 kDa y el complejo tiene en total 350 kDa) la reconstrucción tridimensional del complejo no mostraba ninguna diferencia aparente con las reconstrucciones previas. Este análisis se repitió varias veces, hasta llegar a la conclusión de que el fragmento carboxi-terminal de la proteína Tho2 debía ser flexible y por tanto no es identificable en la reconstrucción del complejo por el método de partículas individuales (explicado en el apartado 4.3). Esta observación es también consistente con el hecho de que la región carboxi-terminal de la proteína la Tho2 esté expuesta al medio. Para confirmar esta teoría, se recurrió a la predicción de estructura secundaria. El resultado mostró que la proteína Tho2 presenta una región altamente desestructurada el extremo carboxi-terminal (Figura 37), lo cuál resulta consistente con los resultados observado en microscopía electrónica.

Estos datos apuntan a que el complejo THO interacciona con los ácidos nucleicos en una región externa al complejo y que dicha interacción requiere la presencia del extremo carboxi-terminal de la proteína Tho2, que es una región flexible. Además, es posible que esta región sea la responsable de mantener la estabilidad genética durante el proceso de transcripción, lo que explicaría la observación de que, en condiciones desfavorables, esta la presencia del extremo carboxi-terminal de la proteína Tho2 se hace tan imprescindible para la supervivencia celular como el complejo THO completo.

5.4. Hpr1 en la biogénesis de RNPm

La identificación de la proteína Hpr1 en el complejo THO con un anticuerpo que reconoce parte de su extremo carboxi-terminal es consistente con la observación de que la muestra Hpr1-TAP posibilitase la purificación del complejo THO. Así, en esta cepa la cola TAP se expresó en el extremo carboxi-terminal de la proteína y quedó expuesta al medio, permitiendo la purificación por el método de purificación en tándem por afinidad (apartado 3.1.1). Estos datos son también consistentes con la observación de que Hpr1 se incorpora al complejo THO de *S. cerevisiae* por el extremo amino-terminal (García-Rubio *et al.*, 2008) (Figura 38).

MSNTEELIQNSIGFLQKTFKALPVSFDSIRHEPLPSSMLHASVLNFEWEPELEKNISAIHD
 RDSLIDIILKRFIIDSMNTNAIEDEEENNLEKGLLNSCIGLDFVYNSRFNRSNPASWGNTF
 FELFSTIIDLLNSPSTFLKFWPYAESRIEWFKMNTSVEPVSLGESNLISYKQPLYEKL RH
 WNDILAKLENNIDILNTVKHYNMKYKLENFLSELLPINEESNFNRSASISALQESDNEWNR
 SARERESNRSSDVIFAADYNFVFYHLICPIEFAFSDLEYKNDVDRSLPLDAILEIEE
 NFYSKIKMNNRTRYSLLEEALNTEYYANYDVMTPKLPVYMKHSNAMKMDRNEFWANLQNIK
 ESSDYTLRPTIMDISLNTTCLYKQLTQEDDDYYRKQFILQLCFTTNLIRNLISSDETR N
 FYKSCYLRENPLSDIDFENLDEVNKKRGLNLCSYICDNRVLFYKIKDPDFYRVIRKLMS
 SDEKFTTAKIDGKFEQNFRISEKIKPPAFDETFFKFTFIKMGKLNINNVWKIPTGLDK
 IEQEVKKPEGVYEEAAQAKWESKISSETSGGEAKDEIIRQWQTLRFLRSRYLDFDFKVNEK
 TGV DGLFEERPKVEALDDSFKEKLLYKINQEHRKKLQDAREYKIGKERKKRALEEEASF
 EREQIKISQRINSASQTEGDELKSEQTPKGEISEENTKIKSSEVSSQDPDSGVAGEFAP
 QNTTAQLENPKTEDNNAATSNISNGSSTQDMK

Figura 38: **Disposición de la secuencia de la proteína Hpr1 de *S. cerevisiae*.** La figura muestra la secuencia completa de la proteína Hpr1 de *S. cerevisiae* en la que se ha resaltado en colores las diferentes regiones de la secuencia. En naranja la región amino-terminal propuesta como punto de interacción de la proteína con el resto de componentes del complejo THO (García-Rubio et al., 2008). En verde la región de interacción de la proteína Hpr1 con el factor de exportación Mex67 (Hobeika et al., 2009). En azul la región de ubiquitinación por la proteína Rsp5 (Gwizdek et al., 2005). En rojo el epítipo que reconoce el anticuerpo anti-Hpr1 utilizado para inmunolocalización de la proteína en el complejo THO.

La proteína Hpr1 es el componente del complejo THO mejor caracterizado hasta la fecha. Si la función principal del complejo es la biogénesis cotranscripcional de RNPm (Aguilera, 2005), Hpr1 es un componente clave para la realización de la función del complejo THO (García-Rubio *et al.*, 2008), dado que es la responsable directa de la interacción del complejo con la ARNP II y con prácticamente todos los factores de exportación con los que interacciona el THO, definidos hasta el momento. Hpr1 fue la primera proteína del complejo THO en ser identificada (Aguilera *et al.*, 1989; 1990) y pronto se describió como asociada con la ARNP II durante la elongación transcripcional (Chang *et al.*, 1999). Esta interacción convierte a Hpr1 en la responsable de que el complejo THO recorra el gen a lo largo de la transcripción para realizar su función (Rondon *et al.*, 2003b). Hpr1 se considera la responsable de la formación y de la degradación del complejo THO y es la única proteína implicada en la biogénesis de RNPm para la que se ha descrito un sitio concreto de regulación y degradación (ubiquitinación) (Gwizdek *et al.*, 2005).

Hpr1 recluta cotranscripcionalmente el factor de exportación Sub2 (Zenklusen *et al.*, 2002), una proteína esencial que adopta un papel primordial en la

maduración del ARNm y la biogénesis de RNPm en todas las células eucariotas (Strasser *et al.*, 2001). También es responsable de la incorporación del factor de exportación Mex67 sobre el ARNm (Gwizdek *et al.*, 2006; Hobeika *et al.*, 2007), acción que determina el final de la transcripción y el principio de la exportación citoplasmática de los RNPm (Hurt *et al.*, 2000). Además, los factores de exportación Gbp2 y Hrb1 dependen de Hpr1 y de Mft1 para su incorporación al ARNm (Hacker *et al.*, 2004; Hurt *et al.*, 2004). Así, la localización de Hpr1 en el complejo THO aporta información muy valiosa para el posicionamiento del complejo THO en la maquinaria de transcripción y ayuda a imaginar la forma en la que tiene lugar la incorporación de factores de transcripción al complejo THO durante su función en la elongación transcripcional (Figura 39).

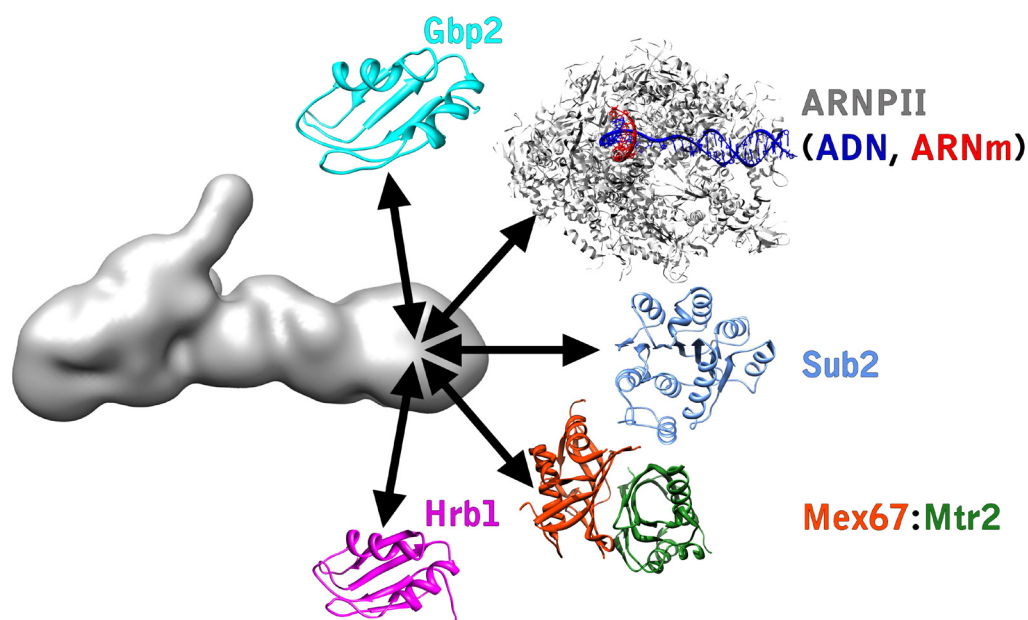


Figura 39: Proteínas que interaccionan con Hpr1 en *S. cerevisiae*. La figura muestra la reconstrucción tridimensional del complejo THO junto con las estructuras y modelos atómicos de las proteínas que interaccionan directamente con el componente Hpr1. Las flechas representan la interacción entre las proteínas y la región en el complejo en la que ha sido localizada Hpr1. El lugar exacto por el que interacciona cada proteína con Hpr1 no ha sido descrito por lo que las flechas son únicamente indicativas de interacción y no contemplan el lugar de Hpr1 por el que ocurre la interacción. En la figura los tamaños de las proteínas no son proporcionales ni consistentes con el tamaño real de cada una de ellas. Las proteínas representadas se enumeran a continuación, de arriba abajo: Gbp2 (modelo atómico computacional que comprende desde el aminoácido 215 al 297 generado por el repositorio automático de Swiss Model), ARNPII (estructura de la ARNPII interaccionando con el ADN, en azul y el ARN, en rojo, resuelta por cristalografía de rayos X y depositada en la base de datos de proteínas PDB con el número de acceso 1Q40), Sub2 (modelo atómico computacional que comprende desde el aminoácido 62 hasta el 270 generado por el repositorio automático de Swiss Model en base a la estructura de su homóloga humana Uap56), Mex67 en complejo con Mtr2 (estructura resuelta por cristalografía de rayos X y depositada en la base de datos de proteínas PDB con el número de acceso 1XTI) y Hrb1 (modelo atómico computacional que comprende desde el aminoácido 156 hasta el 238 generado por el repositorio automático de Swiss Model).

5.5. Expresión génica en *S. cerevisiae*: modelo de acción

En el estudio de los sistemas biológicos, los científicos han diseñado tradicionalmente sus experimentos atendiendo a dos tipos de condiciones, llamadas *in vivo* e *in vitro*. Los experimentos *in vivo* son aquellos realizados en un tejido o dentro de un organismo vivo, mientras que los experimentos *in vitro* se realizan fuera de las condiciones biológicas originales, en un ambiente controlado por el investigador. En los últimos años el avance tecnológico ha permitido adquirir una nueva dimensión en el abordaje del estudio de los sistemas biológicos, definiéndose así los llamados estudios *in silico* (Geris *et al.*, 2010; Ma *et al.*, 2010). Los estudios *in silico* analizan los sistemas biológicos desde el punto de vista computacional, de forma que pueden alcanzarse conclusiones derivadas de cálculos genéricos realizados por un ordenador (Khachaturian *et al.*, 2009; Sleator *et al.*, 2010; Takahashi *et al.*, 2010). Los estudios presentados en esta tesis doctoral incluyen resultados derivados de los tres tipos de abordajes. Así, los estudios fenotípicos se hacen *in vivo*, los estudios de interacción con ácidos nucleicos y la formación de inmunocomplejos se hacen *in vitro* y las reconstrucciones tridimensionales, la predicción de estructuras y la localización de proteínas en los complejos se hacen *in silico*.

Combinando los resultados obtenidos en los diferentes abordajes empleados en este trabajo con los presentados previamente en la bibliografía se consigue una mejor caracterización del proceso de expresión génica. A continuación se presenta un modelo de acción para el desarrollo de las primeras fases de la expresión génica en *S. cerevisiae*, en el que se resalta la posible funcionalidad del complejo THO. El modelo pretende combinar los resultados previos existentes con los obtenidos en esta tesis doctoral, para describir un modelo de acción generalizado mas completo de las fases tempranas de la expresión génica. Al tratarse de un modelo para *S. cerevisiae* y con el objetivo de simplificar la explicación, se ha supuesto la expresión de un gen que no contiene intrones, asumiendo que sólo el 5% de los genes de levaduras presentan intrones (Parenteau *et al.*, 2008).

La expresión génica comenzaría con la activación de genes para la transcripción. La activación de un determinado gen tiene lugar, en general, a través

de la interacción entre unas proteínas conocidas como “factores activadores de la transcripción” y el fragmento de ADN correspondiente (Rafalska-Metcalf *et al.*, 2010). Estas proteínas generalmente actúan desplazando a otras que se encuentran previamente unidas al gen, llamadas “represoras” o “silenciadoras” (Chandler *et al.*, 2001). Con la activación de los genes tiene lugar el inicio de la transcripción (apartado 1.2) que consiste en el acercamiento de una serie de proteínas al gen, que facilitan el acceso del complejo de transcripción al ADN. En esta fase, el fragmento de ADN correspondiente al gen es desplazado hacia la membrana nuclear, de forma que la transcripción sucede cerca del perímetro interior del núcleo (Dieppo *et al.*, 2010). Así, queda definido el “lugar de la transcripción”, que será la ubicación dentro del núcleo en la que tendrá lugar el resto del proceso.

Durante la iniciación de la transcripción acuden al lugar de la transcripción múltiples proteínas, entre las que se encuentran varios factores de transcripción, la enzima ARNP_{II}, (que será el “motor” de la transcripción) y probablemente el complejo THO (o quizás el complejo THO:Tex1) (Santos-Rosa *et al.*, 1994; Rafalska-Metcalf *et al.*, 2010). Los factores de transcripción reconocen la secuencia correspondiente al comienzo del gen (extremo 5' terminal), conocida como secuencia promotor. La interacción de los factores de iniciación con la secuencia promotor desencadena la formación del complejo de transcripción, en el que participa una helicasa que promueve la apertura de la doble hebra de ADN. Así, la ARNP_{II} accede finalmente a la hebra molde de ADN e interacciona con ella dando lugar a la llamada burbuja de transcripción, en la que comienza la síntesis del ARNm como consecuencia de la formación de enlaces fosfodiéster entre ribonucleótidos, favorecida por la ARNP_{II} (apartado 1.2).

En las fases tempranas de la transcripción el complejo THO interacciona con la ARNP_{II}. Esta interacción tiene lugar entre la proteína Hpr1 del complejo THO y el CTD de la ARNP_{II} de forma que cuando comienza la transcripción el complejo THO queda situado detrás de la ARNP_{II} (Figura 40). Probablemente la proteína Tex1 se asocie al complejo THO en una fase temprana de la transcripción y dada su proximidad a la proteína Hpr1, podría funcionar facilitando la interacción entre el complejo THO y la ARNP_{II}. A medida que avanza la transcripción, el

ARNm naciente va saliendo del complejo de transcripción por el CTD, que hace el efecto de “plataforma” para la recepción de los factores que actúan durante la transcripción (Jimeno *et al.*, 2008). Probablemente el complejo THO:Tex1, contribuya también a la formación de esa “plataforma” en la que son recibidos los factores de maduración y exportación del ARNm. Según los resultados obtenidos en el análisis de la localización de la proteína Tex1 y en la predicción de su estructura tridimensional, se podría pensar que el dominio en β -propeller de Tex1 desempeña un papel en la recepción de factores de transcripción y contribuye a la formación de la “plataforma”, facilitando las interacciones de unas proteínas con otras a lo largo de la transcripción. Así, la estructura en β -propeller quedaría localizado cerca de Hpr1 y de la ARNP_{II} y funcionaría de “esqueleto” rígido para facilitar las interacciones proteína-proteína, como se ha descrito para otras proteínas de la familia WD organizadas en estructuras de tipo β -propeller (Smith *et al.*, 1999).

El CTD de la ARNP_{II} presenta una estructura parcialmente desestructurada, al igual que se ha propuesto para el extremo carboxi-terminal de la proteína Tho2 (apartado 5.3). El CTD podría regular el reclutamiento de factores según sus diferentes estados de fosforilación (Masuda *et al.*, 2005) mientras que la región carboxi-terminal de Tho2 podría funcionar evitando la formación de híbridos ADN:ARN y el consecuente daño genético mediante su interacción con las cadenas nucleotídicas de ADN y ARN (Gómez-González *et al.*, 2009b). Así, esta región, podría interaccionar con la hebra complementaria de ADN, que se encuentra previsiblemente libre, protegiéndola del daño genético inducido por la enzima AID (*Activation-Induced Cytidine Deaminase*) (Gómez-González *et al.*, 2007). Como alternativa, la región carboxi-terminal de la proteína Tho2 podría interaccionar con el ARNm recién sintetizado, facilitando la carga de factores de exportación sobre la cadena nucleotídica o protegiéndola de posibles agentes dañinos externos, como la presencia de luz ultravioleta (González-Barrera *et al.*, 2002). Ambos modelos explicarían el hecho de que el complejo THO prevenga la inestabilidad genética asociada a la transcripción y evite la formación de híbridos ADN:ARN a la vez que dirige la biogénesis de RNP_m (Huertas *et al.*, 2003; Wellinger *et al.*, 2006). Sería coherente pensar que el complejo THO contribuye también, directa o indirectamente, a la regulación del

CTD de la ARNP^{II}. De hecho, los factores de exportación Gbp2 y Hrb1 interactúan directamente con el complejo THO e intervienen en la regulación del CTD a través de su interacción con la kinasa Ctk1 (Hurt *et al.*, 2004) y la proteína Tho2 suprime al regulador de la ARNP^{II} Sin4 (West *et al.*, 2000). Es posible que Tex1, como proteína WD40 localizada cerca de la ARNP^{II}, también ejerza un papel en la adaptación de estos factores al complejo THO o al CTD de la ARNP^{II} y contribuya de forma indirecta a la regulación de la transcripción.

Siguiendo con el proceso, la transcripción avanzaría hacia el extremo 3' del gen, a medida que se sintetiza el ARNm. En las primeras fases de transcripción tendría lugar la formación de la caperuza 5' en los nucleótidos iniciales de la cadena de ARNm (apartado 1.3 de Introducción). A medida que la ARNP^{II} recorriese el gen, en complejo THO iría previsiblemente cargando factores de exportación a lo largo de la cadena nucleotídica, favoreciendo la biogénesis cotranscripcional de RNPm. La estructura alargada del complejo THO, con la proteína Hpr1 en uno de sus extremos interactuando con la ARNP^{II}, podría suponer una distribución del eje principal del complejo mas o menos en paralelo a la cadena de ARNm. Así, uno de los extremos del complejo estaría mas cerca de la ARNP^{II} y del ARNm funcionando en la biogénesis de RNPm mientras que el otro extremo quedaría mas alejado, probablemente contribuyendo al mantenimiento de la estabilidad genética transcripcional.

La ARNP^{II} también intervendría en la carga de factores sobre el ARNm recién sintetizado, quizá en una fase previa a la función del complejo THO, dado que éste quedaría localizado detrás de la ARNP^{II}. Npl3 es un factor de exportación esencial cuyo reclutamiento depende exclusivamente de la ARNP^{II} y ocurre en los primeros pasos de la síntesis del ARNm (Krebber *et al.*, 1999; Lei *et al.*, 2001; Lei *et al.*, 2002; Hacker *et al.*, 2004). Por el contrario, Sub2, Yra1, Gbp2 y Hrb1 son algunos de los factores de exportación que dependerían del complejo THO para ser reclutados sobre la cadena de ARNm recién sintetizada. Los componentes del complejo THO que han sido propuestos como responsables de la carga de estos factores son Hpr1 y Mft1 (Strasser *et al.*, 2001; Zenklusen *et al.*, 2002; Hacker *et al.*, 2004). Hpr1 estaría localizada en la zona por la que el ARNm emerge de la ARNP^{II}, por lo que podría colaborar en el reclutamiento de factores de exportación inmediatamente

después de la síntesis de la cadena de ARNm. Además se cree que el reclutamiento de estos factores ocurre en cadena, de forma que unos reclutan a otros, como es el caso de Hpr1 que interacciona con Sub2 y a continuación Sub2 interacciona con Yra1 posicionándola sobre el ARNm (Strasser *et al.*, 2001; Zenklusen *et al.*, 2002). Los factores reclutados se distribuirían a lo largo de la cadena de ARNm interactuando con ella y al mismo tiempo, entre ellos (Hurt *et al.*, 2004).

En algún momento, antes o durante la transcripción, la proteína Hpr1 del complejo THO sería marcada por una cola de poliubiquitinas por acción de la proteína Rsp5 (Gwizdek *et al.*, 2005). De nuevo, la localización estratégica de la proteína Tex1, en una zona central del complejo y próxima a la proteína Hpr1 podría contribuir a la interacción entre Rsp5 y Hpr1. El factor de exportación Mex67 interaccionaría con Hpr1 mediante su dominio carboxi-terminal (dominio de tipo UBA) dejando la cola de poliubiquitinas inaccesible al medio y evitando así la degradación de Hpr1 (Gwizdek *et al.*, 2006). Se ha propuesto que de esta forma la proteína Hpr1 podría regular la exportación de ARNm, a través de la formación del complejo THO y del mantenimiento de su estabilidad (Gwizdek *et al.*, 2005; Hobeika *et al.*, 2007). La propuesta de que Tex1 acompañe al complejo THO durante la elongación transcripcional facilitando la interacción con otros factores es consistente con la observación de que esta proteína se hace mas importante para la expresión génica en condiciones desfavorables (apartado 4.11), donde su acción aumenta la cantidad de transcritos viables que llegan a expresarse como proteínas en el citoplasma celular. Dado que Tex1 es la única proteína asociada al complejo THO que no interacciona directamente con el ADN ni con el ARN, sería lógico pensar que Tex1 no forma parte de los mecanismos de mantenimiento de la estabilidad genética, que serían responsabilidad de la proteína Tho2 (apartado 5.3). De esta forma, si asumimos que la supervivencia celular puede estar condicionada por la inestabilidad genética, concluimos que este modelo es consistente con la observación de que Tex1 no influye en la supervivencia celular (apartado 4.11).

El final de la transcripción llegaría cuando la ARNP_{II} alcanzase el extremo 3' del gen, marcado por la presencia de una secuencia de nucleótidos específica o por unas proteínas que interaccionan específicamente con el ADN al final del gen. Al

transcribir esta secuencia o encontrar dichas proteínas, la ARNPII se desestabilizaría y provocaría un cambio en la maquinaria de transcripción que facilitaría la intervención de la maquinaria responsable del corte y poliadenilación del extremo 3' del ARNm. Así, esta maquinaria añadiría un tramo de secuencia al ARNm compuesto por bases de adenina que protegería al ARNm de la degradación citoplasmática. Se ha propuesto que el factor de exportación Sub2 contribuye a la formación de la cola de poliadenilación (Rougemaille *et al.*, 2008) y se ha observado que la adición de la cola de poliadenilación es imprescindible para el reclutamiento del factor de exportación Yra1 y para la exportación del ARNm (Lei *et al.*, 2002). Después del corte en el extremo 3', el ARNm sería liberado del lugar de la transcripción (Figura 1) e inmediatamente transportado hacia el complejo de poros de la membrana nuclear (NPC), que se encontraría relativamente próximo al lugar de la transcripción. Se ha propuesto que en los últimos pasos de la transcripción la proteína Mex67 desplaza a Sub2 mediante su interacción con Yra1, de forma que Mex67 es cargada sobre el ARNm por su interacción con Yra1 y Sub2 es liberada del RNPm (Strasser *et al.*, 2001; Zenklusen *et al.*, 2002). Mex67 presenta baja afinidad por el ARN e interacciona con él por medio de Yra1, que actúa como proteína adaptadora (Strasser *et al.*, 2000a; Gwizdek *et al.*, 2005). Así, Sub2 formaría parte del RNPm durante la primeras fases de la transcripción y después Mex67 la desplazaría con el objetivo de facilitar la exportación de el RNPm. Esta hipótesis tiene sentido dado que la proteína Sub2 es además un factor de *splicing* y su función sería mas importante en la primera fase de la transcripción (Jensen *et al.*, 2001b).

En el momento en que el ARNm fuese liberado del lugar de la transcripción, toda la molécula se encontraría rodeada de proteínas que interaccionarían unas con otras definiendo una estructura compacta conocida como RNPm. Algunas de las proteínas que componen el RNPm interaccionarían con los componentes del NPC, facilitando su translocación hacia el citoplasma celular (Stewart, 2007). El factor de exportación Mex67 interacciona directamente con la proteína exportadora Mtr2 que a su vez interacciona con las proteínas del canal central del NPC, favoreciendo el ajuste del RNPm en el interior del NPC (Strasser *et al.*, 2000b; Strasser *et al.*, 2000a; Gwizdek *et al.*, 2006). Los factores Npl3 y Nab2 también formarían parte del RNPm

y contribuirían a la translocación nucleocitoplasmática del RNPm a través de su interacción específica con el NPC (Krebber *et al.*, 1999; Green *et al.*, 2003; Marfatia *et al.*, 2003). Algunos de los componentes del RNPm acompañarían al ARNm hasta el ribosoma mientras que otros serían importados al núcleo después de la exportación (Rodríguez *et al.*, 2004). Una vez exportada el RNPm, las proteínas que lo componen determinan la localización, la traducción y la estabilidad del ARNm en el citoplasma (Dreyfuss *et al.*, 2002) (Figura 40).

Durante todo el proceso de transcripción, existen varios mecanismos que verifican la funcionalidad del ARNm, conocidos como mecanismos de control de calidad del ARNm. Algunos de estos mecanismos se encuentran asociados a las fases iniciales de la biogénesis de RNPm, de forma que los ARNm defectuosos son retenidos y degradados por el exosoma nuclear cotranscripcionalmente (Olesen *et al.*, 2005). Se ha descrito que el complejo THO interacciona directamente con las maquinarias de NER (*Nucleotide Excision Repair*) y TCR (*Transcription Coupled Repair*) (Selby *et al.*, 1990; González-Barrera *et al.*, 2002) y participa activamente en la regulación de estos procesos. Así, en el modelo de acción habría que incluir las maquinarias de control de calidad y el exosoma nuclear, que se encargaría de eliminar las moléculas de ARNm defectuosas.

Cabría decir que la transcripción y la biogénesis de RNPm no ocurren de igual forma en todos los genes. Se ha comprobado experimentalmente que la función del complejo THO es especialmente importante para la transcripción de genes que requieren una elevada tasa de expresión (Jimeno *et al.*, 2008). Así, mientras que una tasa baja de transcripción facilita el reclutamiento de factores de exportación en ausencia del complejo THO, la transcripción de genes que requieren una alta tasa de expresión en un momento determinado, como el caso de los genes de respuesta al estrés térmico, es altamente dependiente de la función del complejo THO (Hurt *et al.*, 2000; Jensen *et al.*, 2001a). Este hecho es consistente además con la observación de que las altas tasas de transcripción aumentan la inestabilidad genómica, lo que hace mas importante la acción del complejo THO (Olesen *et al.*, 2005). Además, se ha propuesto que la función del complejo THO se hace tanto mas necesaria cuanto mayor número de repeticiones presente el gen a transcribir (Rondon *et al.*, 2003a;

Voynov *et al.*, 2006). Aún con todo, en los últimos años se han descrito mas de 400 genes de levadura que requieren la acción del complejo THO para su expresión (Rougemaille *et al.*, 2008).

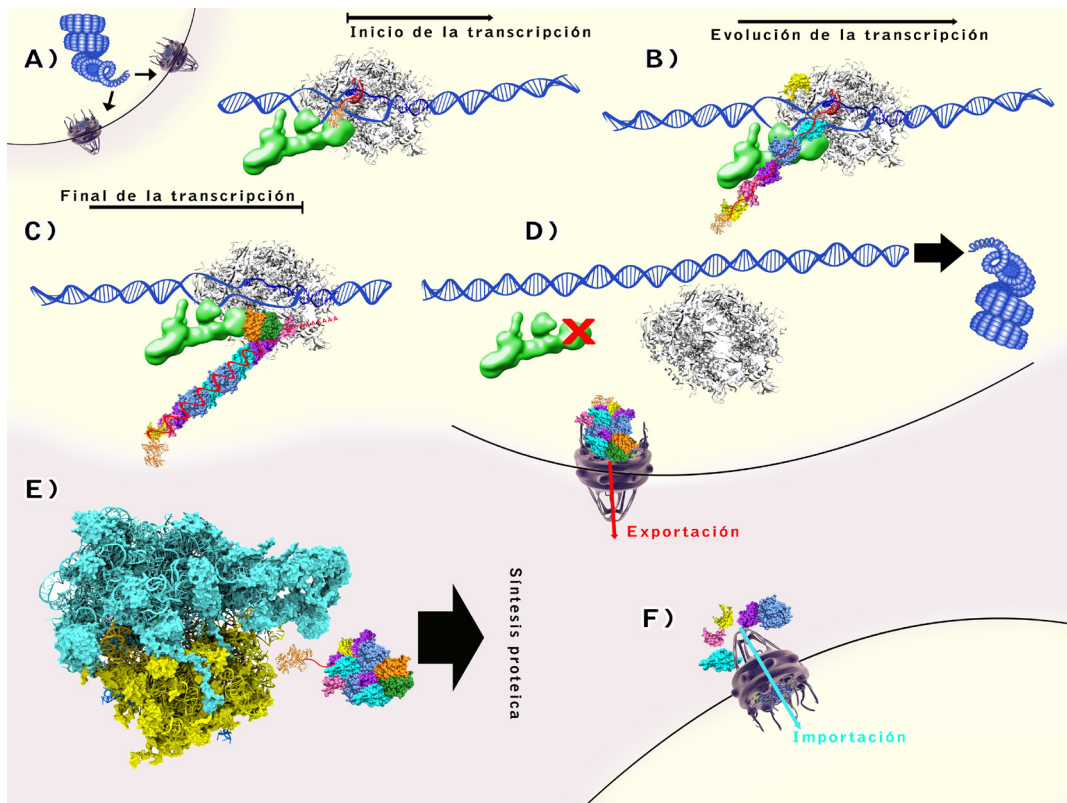
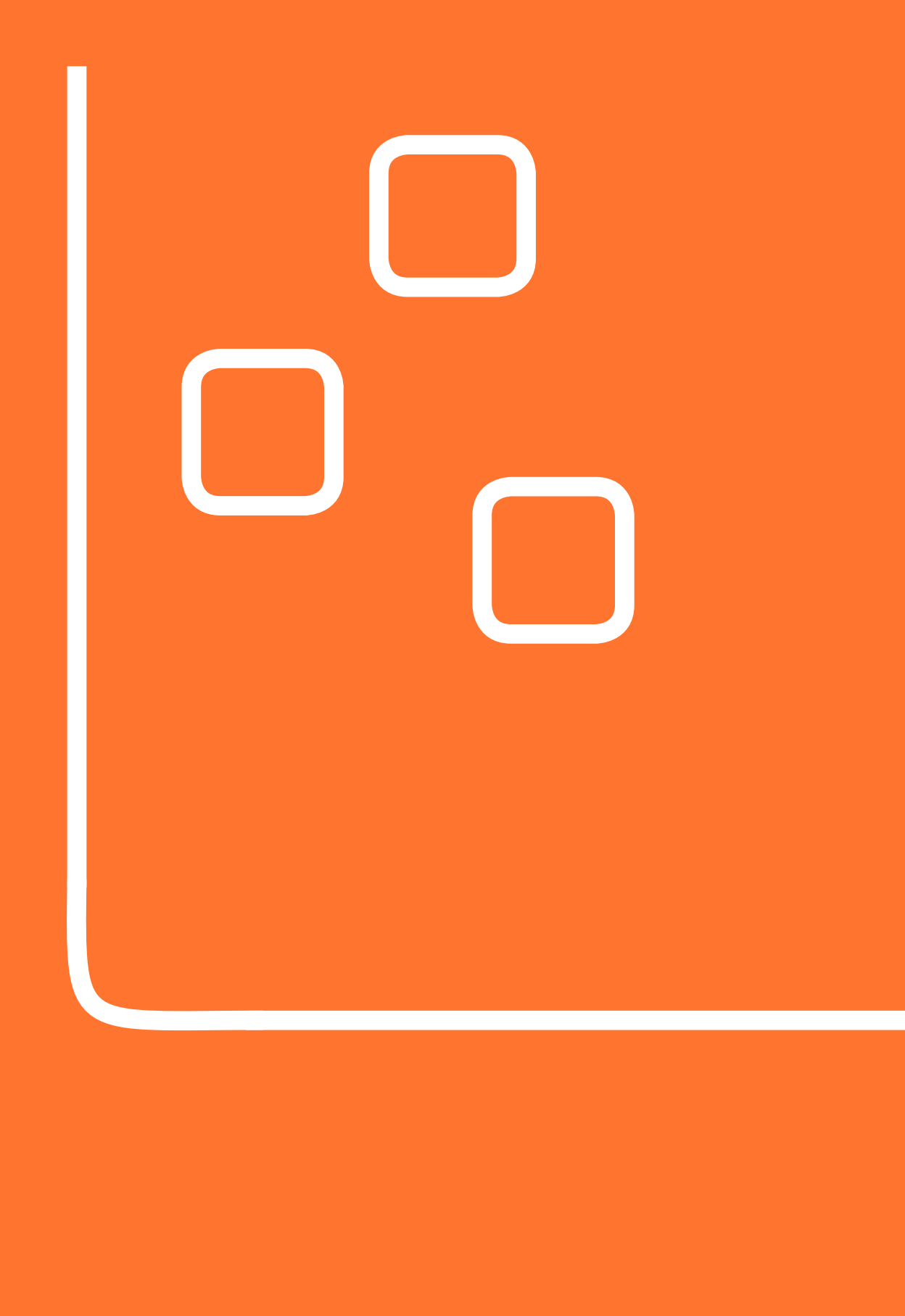


Figura 40: Modelo esquemático simplificado de un posible ciclo de expresión génica en *S. cerevisiae*: funcionamiento del complejo THO. **A)** Inicio de la transcripción. Durante la activación, el gen que va a ser transcrito sería desplazado hacia la membrana nuclear (línea negra) cerca del NPC (estructuras moradas) y allí comenzaría la transcripción con la unión de la ARNPII (estructura de color gris) y probablemente del complejo THO:Tex1 (en color verde) a la molécula de ADN (en azul). Con la transcripción empezaría la síntesis del ARNm (en rojo) con el correspondiente marcaje de la caperuza en el extremo 5' (estructura en naranja). **B)** Elongación transcripcional. A medida que el ARNm fuese siendo sintetizado, sería simultáneamente recubierto por varios factores de exportación (a lo largo de la cadena de ARNm). Tanto la ARNPII como el complejo THO participarían en el proceso de biogénesis cotranscripcional de RNPm, de forma que unos factores serían reclutados sobre el ARNm por la ARNPII (Npl3 en amarillo) y otros por el complejo THO (Sub2, Yra1, Gbp2 y Hrb1 en azul, morado, cian y rosa respectivamente). **C)** Final de la transcripción. Una sintetizado el ARNm, el complejo THO favorecería la interacción del dímero Mex67:Mtr2 (en naranja y verde) sobre el ARNm y la formación de la cola de poliadenilación del extremo 3' (letras A rojas). Finalmente el ARNm se soltaría de la ARNPII y del ADN. **D)** Exportación citoplasmática. El complejo THO sería degradado empezando por la proteína Hpr1 (representado por una cruz roja) y el ADN volvería a ser almacenado en una forma compacta (representado por una flecha negra). El ARNm protegido por proteínas (RNPm) atravesaría NPC, favorecido por la interacción del dímero Mex67:Mtr2 con el NPC. **E)** Traducción. Una vez en el citoplasma el RNPm alcanzaría el ribosoma y el ARNm sería liberado para ser "leído" por el ribosoma par la síntesis proteica. **F)** Importación nuclear. Una vez sintetizadas las proteínas el ARNm sería probablemente degradado en el citoplasma y algunos de los facotres de exportación que componían el RNPm seían importados de nuevo al núcleo a través del NPC (diferentes factores siguen distintas rutas de importación). En la figura el fondo amarillo corresponde al espacio dentro del nucleo y el rojizo al espacio citoplasmático. **DATOS:** factores de exportación: *Sub2* (azul), *Yra1* (morado), *Hrb1* (rosa), *Gbp2* (cian), *Npl3* (amarillo) y *Mex67:Mtr2* (naranja y verde). ARNPII (estructua 3M3Y del PDB), caperuza 5' (1N52 del PDB), ribosoma y NPC (www.computerweekly.com). Las proteínas en cursiva han sido obtenidas a prtir de modelos paricales obtenidos en el repositorio de Swiss-Model.

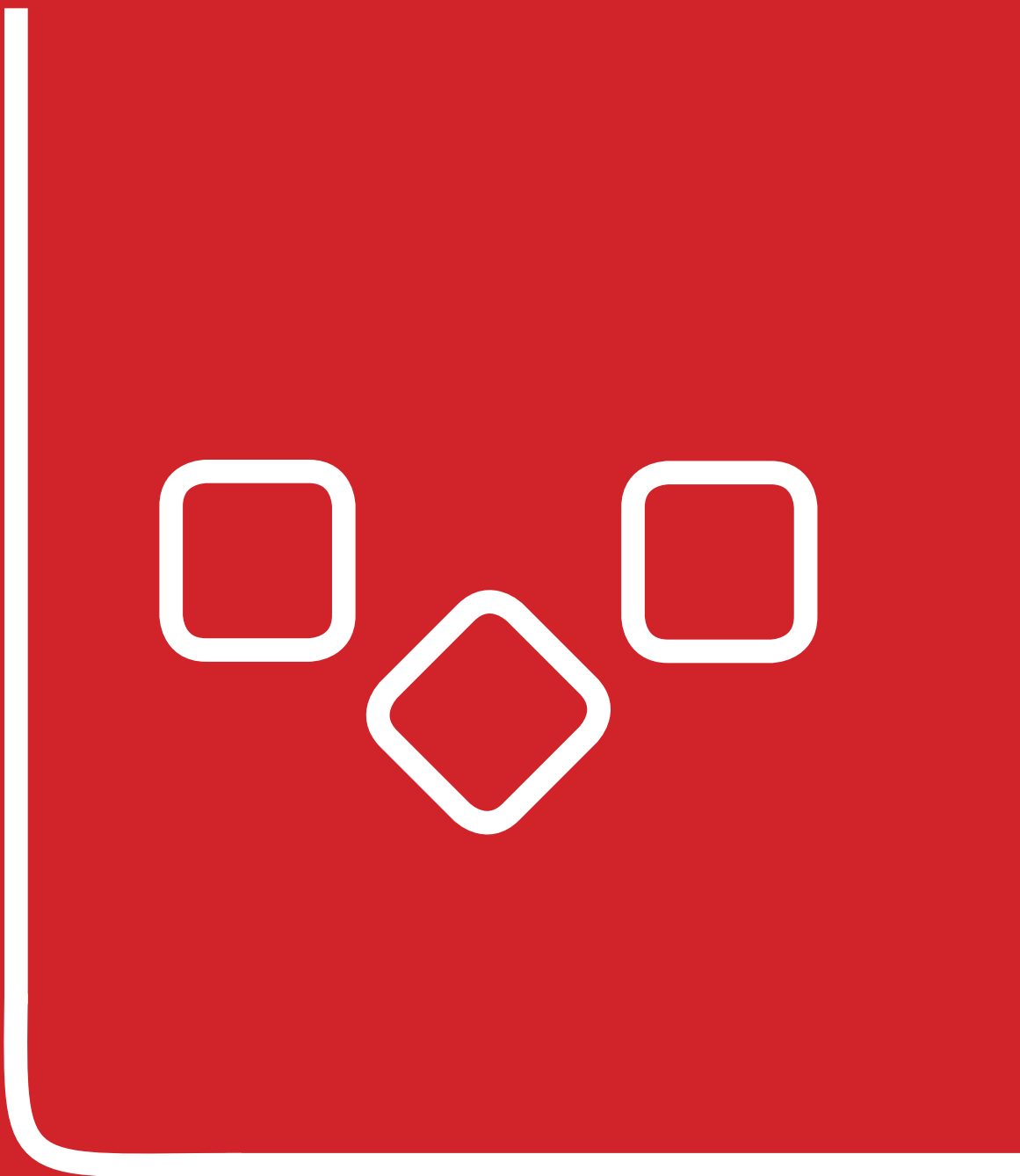
Los resultados presentados en esta tesis doctoral aportan información novedosa sobre la estructura del complejo THO de *S. cerevisiae*, su relación con la proteína WD Tex1 y sobre su interacción con los ácidos nucleicos. Sin embargo, en el objetivo de comprender el funcionamiento y la evolución de las primeras fases de la expresión génica, aún queda bastante trabajo por hacer. Determinar la localización de los cuatro componentes en la reconstrucción del complejo THO y definir así la organización del complejo podría ayudar a comprender la función que desempeña cada proteína. El análisis estructural de otros complejos que involucren al complejo THO junto con los ácidos nucleicos y otras proteínas relacionadas proporcionaría información útil para describir con mayor exactitud el transcurso del proceso de transcripción. Sería interesante también caracterizar estructuralmente los complejos THO de otros organismos y compararlos con el de *S. cerevisiae* para tener así un registro de la evolución estructural del complejo, en los diferentes organismos. Desde el punto de vista bioquímico se podría continuar con la caracterización fenotípica de las cepas mutantes en el complejo THO y estudiar su comportamiento en otras condiciones, por ejemplo estrés salino, supervivencia en bajas temperaturas o expresión de diferentes genes. Quizá estas y otras aproximaciones llevarían a una mejor comprensión del proceso de expresión génica y al diseño de tratamientos para algunas de las enfermedades en las que se ha visto implicado el complejo THO de humanos.





6. Conclusiones

- 1.- El complejo THO de *S. cerevisiae*, formado por las proteínas Tho2, Hpr1, Mft1 y Thp2 copurifica con la proteína Tex1, formando el complejo THO:Tex1.
- 2.- La reconstrucción tridimensional del complejo THO:Tex1 presenta una estructura alargada en la que la proteína Hpr1 se localiza en uno de sus extremos mientras que Tex1 lo hace por la zona central.
- 3.- La predicción computacional de estructura de la proteína Tex1 muestra una región amino-terminal con un plegamiento de tipo β -*propeller* que se ajusta correctamente en el volumen correspondiente a Tex1 en la reconstrucción tridimensional del complejo THO:Tex1.
- 4.- El complejo THO de *S. cerevisiae* interacciona con el ADN de cadena sencilla, el ADN de cadena doble y el ARN a través del extremo carboxi-terminal de la proteína Tho2. Este dominio desempeña un papel importante en el desarrollo de la función fisiológica del complejo THO y contribuye a la supervivencia celular.





7. Bibliografía

- Afasizhev R. 2007. [RNA editing]. *Mol Biol (Mosk)* 41: 260-273.
- Aguilera A. 2005. Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Curr Opin Cell Biol* 17: 242-250.
- Aguilera A, Klein HL. 1988. Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae*. I. Isolation and genetic characterization of hyper-recombination mutations. *Genetics* 119: 779-790.
- Aguilera A, Klein HL. 1989. Genetic and molecular analysis of recombination events in *Saccharomyces cerevisiae* occurring in the presence of the hyper-recombination mutation *hpr1*. *Genetics* 122: 503-517.
- Aguilera A, Klein HL. 1990. HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the *Saccharomyces cerevisiae* TOP1 gene. *Mol Cell Biol* 10: 1439-1451.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215: 403-410.
- Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.
- Ast G. 2004. How did alternative splicing evolve? *Nat Rev Genet* 5: 773-782.
- Beilharz T, Beddoe T, Landl K, Cartwright P, Lithgow T. 1997. The protein encoded by the MFT1 gene is a targeting factor for mitochondrial precursor proteins, and not a core ribosomal protein. *FEBS Lett* 407: 220-224.
- Bottcher B, Wynne SA, Crowther RA. 1997a. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 386: 88-91.
- Bottcher B, Kiselev NA, Stel'Mashchuk VY, Perevozchikova NA, Borisov AV, Crowther RA. 1997b. Three-dimensional structure of infectious bursal disease virus determined by electron cryomicroscopy. *J Virol* 71: 325-330.
- Bowie JU, Luthy R, Eisenberg D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253: 164-170.
- Boyne JR, Colgan KJ, Whitehouse A. 2008. Recruitment of the complete hTREX complex is required for Kaposi's sarcoma-associated herpesvirus intronless mRNA nuclear export and virus replication. *PLoS Pathog* 4: e1000194.
- Burckstummer T, Bennett KL, Preradovic A, Schutze G, Hantschel O, Superti-Furga G, Bauch A. 2006. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat Methods* 3: 1013-1019.
- Butler JS. 2002. The yin and yang of the exosome. *Trends Cell Biol* 12: 90-96.
- Cartwright P, Beilharz T, Hansen P, Garrett J, Lithgow T. 1997. Mft52, an acid-bristle protein in the cytosol that delivers precursor proteins to yeast mitochondria. *J Biol Chem* 272: 5320-5325.
- Cech TR. 1983. RNA splicing: three themes with variations. *Cell* 34: 713-716.

- Cerna D, Wilson DK. 2005. The structure of Sif2p, a WD repeat protein functioning in the SET3 corepressor complex. *J Mol Biol* 351: 923-935.
- Chandler VL, Vaucheret H. 2001. Gene activation and gene silencing. *Plant Physiol* 125: 145-148.
- Chang M, French-Cornay D, Fan HY, Klein H, Denis CL, Jaehning JA. 1999. A complex containing RNA polymerase II, Paf1p, Cdc73p, Hpr1p, and Ccr4p plays a role in protein kinase C signaling. *Mol Cell Biol* 19: 1056-1067.
- Chávez S, Aguilera A. 1997. The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes Dev* 11: 3459-3470.
- Chávez S, Garcia-Rubio M, Prado F, Aguilera A. 2001. Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in *Saccharomyces cerevisiae*. *Mol Cell Biol* 21: 7054-7064.
- Chávez S, Beilharz T, Rondon AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, Aguilera A. 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *EMBO J* 19: 5824-5834.
- Cheng H, Dufu K, Lee CS, Hsu JL, Dias A, Reed R. 2006. Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* 127: 1389-1400.
- Cole C, Barber JD, Barton GJ. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36: W197-201.
- Cole MD, Cowling VH. 2009. Specific regulation of mRNA cap methylation by the c-Myc and E2F1 transcription factors. *Oncogene* 28: 1169-1175.
- Corry GN, Underhill DA. 2005. Subnuclear compartmentalization of sequence-specific transcription factors and regulation of eukaryotic gene expression. *Biochem Cell Biol* 83: 535-547.
- Cowling VH. 2010. Regulation of mRNA cap methylation. *Biochem J* 425: 295-302.
- Cuéllar J, Martin-Benito J, Scheres SH, Sousa R, Moro F, Lopez-Vinas E, Gomez-Puertas P, Muga A, Carrascosa JL, Valpuesta JM. 2008. The structure of CCT-Hsc70 NBD suggests a mechanism for Hsp70 delivery of substrates to the chaperonin. *Nat Struct Mol Biol* 15: 858-864.
- Custodio N, Carvalho C, Condado I, Antoniou M, Blencowe BJ, Carmo-Fonseca M. 2004. In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. *RNA* 10: 622-633.
- Daffertshofer A, Lamoth CJ, Meijer OG, Beek PJ. 2004. PCA in studying coordination and variability: a tutorial. *Clin Biomech (Bristol, Avon)* 19: 415-428.
- De Silva H, Lee K, Osley MA. 1998. Functional dissection of yeast Hir1p, a WD repeat-containing transcriptional corepressor. *Genetics* 148: 657-667.
- DeLano WL. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Car-

los, CA, USA, <http://www.pymol.org>.

- Dever TE. 1999. Translation initiation: adept at adapting. *Trends Biochem Sci* 24: 398-403.
- Dever TE. 2002. Gene-specific regulation by general translation factors. *Cell* 108: 545-556.
- Diepkins G, Stutz F. 2010. Connecting the transcription site to the nuclear pore: a multi-tether process that regulates gene expression. *J Cell Sci* 123: 1989-1999.
- Dreyfuss G, Kim VN, Kataoka N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* 3: 195-205.
- Durfee T, Mancini MA, Jones D, Elledge SJ, Lee WH. 1994. The amino-terminal region of the retinoblastoma gene product binds a novel nuclear matrix protein that co-localizes to centers for RNA processing. *J Cell Biol* 127: 609-622.
- Edgar RC. 2004a. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792-1797.
- Edgar RC. 2004b. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
- Erkmann JA, Kutay U. 2004. Nuclear export of mRNA: from the site of transcription to the cytoplasm. *Exp Cell Res* 296: 12-20.
- Fan HY, Merker RJ, Klein HL. 2001. High-copy-number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability. *Mol Cell Biol* 21: 5459-5470.
- Flint SJ. 1981. Splicing and the regulation of viral gene expression. *Curr Top Microbiol Immunol* 93: 47-79.
- Frank J. 2006. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. New York: Oxford University Press.
- Frank J. 2009. Single-particle reconstruction of biological macromolecules in electron microscopy--30 years. *Q Rev Biophys* 42: 139-158.
- Frank J, Penczek P, Liu W. 1992. Alignment, classification, and three-dimensional reconstruction of single particles embedded in ice. *Scanning Microsc Suppl* 6: 11-20; discussion 20-12.
- Frank J, Radermacher M, Penczek P, Zhu J, Li Y, Ladjadj M, Leith A. 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* 116: 190-199.
- Fulop V, Jones DT. 1999. Beta propellers: structural rigidity and functional diversity. *Curr Opin Struct Biol* 9: 715-721.
- García-Rubio M, Chavez S, Huertas P, Tous C, Jimeno S, Luna R, Aguilera A. 2008. Different physiological relevance of yeast THO/TREX subunits in gene expression and genome integrity. *Mol Genet Genomics* 279: 123-132.
- Garrett JM, Singh KK, Vonder Haar RA, Emr SD. 1991. Mitochondrial protein import: isolation and characterization of the *Saccharomyces cerevisiae* MFT1 gene. *Mol Gen Genet*

225: 483-491.

- Geris L, Schugart R, Van Oosterwyck H. 2010. In silico design of treatment strategies in wound healing and bone fracture healing. *Philos Transact A Math Phys Eng Sci* 368: 2683-2706.
- Goddard TD, Huang CC, Ferrin TE. 2005. Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. *Structure* 13: 473-482.
- Goddard TD, Huang CC, Ferrin TE. 2007. Visualizing density maps with UCSF Chimera. *J Struct Biol* 157: 281-287.
- Gómez-González B, Aguilera A. 2007. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc Natl Acad Sci U S A* 104: 8409-8414.
- Gómez-González B, Aguilera A. 2009a. R-loops do not accumulate in transcription-defective hpr1-101 mutants: implications for the functional role of THO/TREX. *Nucleic Acids Res* 37: 4315-4321.
- Gómez-González B, Felipe-Abrio I, Aguilera A. 2009b. The S-phase checkpoint is required to respond to R-loops accumulated in THO mutants. *Mol Cell Biol* 29: 5203-5213.
- Gómez-Moreno C, Sancho J. 2003. *ESTRUCTURA DE PROTEINAS*. Barcelona: Ariel Ciencia.
- González-Barrera S, Prado F, Verhage R, Brouwer J, Aguilera A. 2002. Defective nucleotide excision repair in yeast hpr1 and tho2 mutants. *Nucleic Acids Res* 30: 2193-2201.
- Green DM, Johnson CP, Hagan H, Corbett AH. 2003. The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. *Proc Natl Acad Sci U S A* 100: 1010-1015.
- Grudzien E, Kalek M, Jemielity J, Darzynkiewicz E, Rhoads RE. 2006. Differential inhibition of mRNA degradation pathways by novel cap analogs. *J Biol Chem* 281: 1857-1867.
- Guex N, Peitsch MC. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18: 2714-2723.
- Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30 Suppl 1: S162-173.
- Guo S, Hakimi MA, Baillat D, Chen X, Farber MJ, Klein-Szanto AJ, Cooch NS, Godwin AK, Shiekhhattar R. 2005. Linking transcriptional elongation and messenger RNA export to metastatic breast cancers. *Cancer Res* 65: 3011-3016.
- Gwizdek C, Hobeika M, Kus B, Ossareh-Nazari B, Dargemont C, Rodriguez MS. 2005. The mRNA nuclear export factor Hpr1 is regulated by Rsp5-mediated ubiquitylation. *J Biol Chem* 280: 13401-13405.
- Gwizdek C, Iglesias N, Rodriguez MS, Ossareh-Nazari B, Hobeika M, Divita G, Stutz F, Dargemont C. 2006. Ubiquitin-associated domain of Mex67 synchronizes recruitment of

the mRNA export machinery with transcription. *Proc Natl Acad Sci U S A* 103: 16376-16381.

- Hacker S, Krebber H. 2004. Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. *J Biol Chem* 279: 5049-5052.
- Hamada T, Igarashi H, Taguchi R, Fujiwara M, Fukao Y, Shimmen T, Yokota E, Sonobe S. 2009. The putative RNA-processing protein, THO2, is a microtubule-associated protein in tobacco. *Plant Cell Physiol* 50: 801-811.
- Hammell CM, Gross S, Zenklusen D, Heath CV, Stutz F, Moore C, Cole CN. 2002. Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol* 22: 6441-6457.
- Hautbergue GM, Hung ML, Golovanov AP, Lian LY, Wilson SA. 2008. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proc Natl Acad Sci U S A* 105: 5154-5159.
- Hayashi K, Ohba Y. 1972. [SDS-polyacrylamide gel electrophoresis]. *Tanpakushitsu Kakusan Koso* 17: 304-311.
- Hearst JE, Botchan M. 1970. The eukaryotic chromosome. *Annu Rev Biochem* 39: 151-182.
- Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E, Downing KH. 1990. An atomic model for the structure of bacteriorhodopsin. *Biochem Soc Trans* 18: 844.
- Heymann JB. 2001. Bsoft: image and molecular processing in electron microscopy. *J Struct Biol* 133: 156-169.
- Heymann JB, Belnap DM. 2007. Bsoft: image processing and molecular modeling for electron microscopy. *J Struct Biol* 157: 3-18.
- Hieronymus H, Silver PA. 2003. Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. *Nat Genet* 33: 155-161.
- Hieronymus H, Silver PA. 2004a. A systems view of mRNP biology. *Genes Dev* 18: 2845-2860.
- Hieronymus H, Yu MC, Silver PA. 2004b. Genome-wide mRNA surveillance is coupled to mRNA export. *Genes Dev* 18: 2652-2662.
- Hilleren P, McCarthy T, Rosbash M, Parker R, Jensen TH. 2001. Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature* 413: 538-542.
- Hobeika M, Brockmann C, Gruessing F, Neuhaus D, Divita G, Stewart M, Dargemont C. 2009. Structural requirements for the ubiquitin-associated domain of the mRNA export factor Mex67 to bind its specific targets, the transcription elongation THO complex component Hpr1 and nucleoporin FXFG repeats. *J Biol Chem* 284: 17575-17583.
- Hobeika M, Brockmann C, Iglesias N, Gwizdek C, Neuhaus D, Stutz F, Stewart M, Divita G, Dargemont C. 2007. Coordination of Hpr1 and ubiquitin binding by the UBA domain of the mRNA export factor Mex67. *Mol Biol Cell* 18: 2561-2568.
- Huang Y, Carmichael GG. 1996. Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Mol Cell Biol* 16: 1534-1542.

- Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* 12: 711-721.
- Huertas P, Garcia-Rubio ML, Wellinger RE, Luna R, Aguilera A. 2006. An hpr1 point mutation that impairs transcription and mRNP biogenesis without increasing recombination. *Mol Cell Biol* 26: 7451-7465.
- Hurt E, Luo MJ, Rother S, Reed R, Strasser K. 2004. Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. *Proc Natl Acad Sci U S A* 101: 1858-1862.
- Hurt E, Strasser K, Segref A, Bailer S, Schlaich N, Presutti C, Tollervey D, Jansen R. 2000. Mex67p mediates nuclear export of a variety of RNA polymerase II transcripts. *J Biol Chem* 275: 8361-8368.
- Imamoto F. 1973. Translation and transcription of the tryptophan operon. *Prog Nucleic Acid Res Mol Biol* 13: 339-407.
- Jarmolowski A, Boelens WC, Izaurrealde E, Mattaj IW. 1994. Nuclear export of different classes of RNA is mediated by specific factors. *J Cell Biol* 124: 627-635.
- Jeffares DC, Poole AM, Penny D. 1998. Relics from the RNA world. *J Mol Evol* 46: 18-36.
- Jensen TH, Patricio K, McCarthy T, Rosbash M. 2001a. A block to mRNA nuclear export in *S. cerevisiae* leads to hyperadenylation of transcripts that accumulate at the site of transcription. *Mol Cell* 7: 887-898.
- Jensen TH, Boulay J, Rosbash M, Libri D. 2001b. The DECD box putative ATPase Sub2p is an early mRNA export factor. *Curr Biol* 11: 1711-1715.
- Jiang W, Ludtke SJ. 2005. Electron cryomicroscopy of single particles at subnanometer resolution. *Curr Opin Struct Biol* 15: 571-577.
- Jimeno S, Aguilera A. 2010. The THO complex as a key mRNP biogenesis factor in development and cell differentiation. *J Biol* 9: 6.
- Jimeno S, Rondon AG, Luna R, Aguilera A. 2002. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *EMBO J* 21: 3526-3535.
- Jimeno S, Garcia-Rubio M, Luna R, Aguilera A. 2008. A reduction in RNA polymerase II initiation rate suppresses hyper-recombination and transcription-elongation impairment of THO mutants. *Mol Genet Genomics* 280: 327-336.
- Jones DT, Taylor WR, Thornton JM. 1992. A new approach to protein fold recognition. *Nature* 358: 86-89.
- Jones DT, Tress M, Bryson K, Hadley C. 1999. Successful recognition of protein folds using threading methods biased by sequence similarity and predicted secondary structure. *Proteins Suppl* 3: 104-111.
- Jukes TH. 1963. Some Recent Advances in Studies of the Transcription of the Genetic Mes-

sage. *Adv Biol Med Phys* 9: 1-41.

- Jurica MS, Moore MJ. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 12: 5-14.
- Kadowaki T, Hitomi M, Chen S, Tartakoff AM. 1994. Nuclear mRNA accumulation causes nucleolar fragmentation in yeast mtr2 mutant. *Mol Biol Cell* 5: 1253-1263.
- Kangas JA, Kohonen TK, Laaksonen JT. 1990. Variants of self-organizing maps. *IEEE Trans Neural Netw* 1: 93-99.
- Kataoka N, Yong J, Kim VN, Velazquez F, Perkinson RA, Wang F, Dreyfuss G. 2000. Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* 6: 673-682.
- Keene JD. 2003. Organizing mRNA export. *Nat Genet* 33: 111-112.
- Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4: 363-371.
- Khachaturian ZS, Lombardo J. 2009. In silico modeling system: a national research resource for simulation of complex brain disorders. *Alzheimers Dement* 5: 1-4.
- Kim M, Ahn SH, Krogan NJ, Greenblatt JF, Buratowski S. 2004. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J* 23: 354-364.
- Kohonen T. 1990. Cortical maps. *Nature* 346: 24.
- Kramer A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* 65: 367-409.
- Krebber H, Taura T, Lee MS, Silver PA. 1999. Uncoupling of the hnRNP Npl3p from mRNAs during the stress-induced block in mRNA export. *Genes Dev* 13: 1994-2004.
- Kubota H. 2009. Quality control against misfolded proteins in the cytosol: a network for cell survival. *J Biochem* 146: 609-616.
- Kumar A, Lindberg U. 1972. Characterization of messenger ribonucleoprotein and messenger RNA from KB cells. *Proc Natl Acad Sci U S A* 69: 681-685.
- Lamond AI. 1993. The spliceosome. *Bioessays* 15: 595-603.
- Lange CF. 1970. Advances in electrophoretic and chromatographic technics. *Prog Clin Pathol* 3: 130-162.
- Lei EP, Silver PA. 2002. Intron status and 3'-end formation control cotranscriptional export of mRNA. *Genes Dev* 16: 2761-2766.
- Lei EP, Krebber H, Silver PA. 2001. Messenger RNAs are recruited for nuclear export during transcription. *Genes Dev* 15: 1771-1782.
- Li D, Roberts R. 2001. WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cell Mol Life Sci* 58: 2085-2097.
- Li Y, Wang X, Zhang X, Goodrich DW. 2005. Human hHpr1/p84/Thoc1 regulates transcriptional elongation and physically links RNA polymerase II and RNA processing factors. *Mol Cell Biol* 25: 4023-4033.

- Li Y, Lin AW, Zhang X, Wang Y, Wang X, Goodrich DW. 2007. Cancer cells and normal cells differ in their requirements for Thoc1. *Cancer Res* 67: 6657-6664.
- Libri D, Dower K, Boulay J, Thomsen R, Rosbash M, Jensen TH. 2002. Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Mol Cell Biol* 22: 8254-8266.
- Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD. 2008. The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* 15: 819-826.
- Llorca O. 2005. Introduction to 3D reconstruction of macromolecules using single particle electron microscopy. *Acta Pharmacol Sin* 26: 1153-1164.
- Ludtke SJ, Baldwin PR, Chiu W. 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128: 82-97.
- Luna R, Gaillard H, Gonzalez-Aguilera C, Aguilera A. 2008. Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. *Chromosoma* 117: 319-331.
- Luna R, Jimeno S, Marin M, Huertas P, Garcia-Rubio M, Aguilera A. 2005. Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Mol Cell* 18: 711-722.
- Luo MJ, Reed R. 1999. Splicing is required for rapid and efficient mRNA export in metazoans. *Proc Natl Acad Sci U S A* 96: 14937-14942.
- Luque D, Rivas G, Alfonso C, Carrascosa JL, Rodriguez JF, Caston JR. 2009. Infectious bursal disease virus is an icosahedral polyploid dsRNA virus. *Proc Natl Acad Sci U S A* 106: 2148-2152.
- Ma XH, Shi Z, Tan C, Jiang Y, Go ML, Low BC, Chen YZ. 2010. In-silico approaches to multi-target drug discovery : computer aided multi-target drug design, multi-target virtual screening. *Pharm Res* 27: 739-749.
- Macgillivray AJ, Paul J, Threlfall G. 1972. Transcriptional regulation in eukaryotic cells. *Adv Cancer Res* 15: 93-162.
- Mancini A, Niemann-Seyde SC, Pankow R, El Bounkari O, Klebba-Farber S, Koch A, Jaworska E, Spooncer E, Gruber AD, Whetton AD, Tamura T. 2010. THOC5/FMIP, an mRNA export TREX complex protein, is essential for hematopoietic primitive cell survival in vivo. *BMC Biol* 8: 1.
- Marabini R, Masegosa IM, San Martin MC, Marco S, Fernandez JJ, de la Fraga LG, Vaquerizo C, Carazo JM. 1996. Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol* 116: 237-240.
- Marfatia KA, Crafton EB, Green DM, Corbett AH. 2003. Domain analysis of the *Saccharomyces cerevisiae* heterogeneous nuclear ribonucleoprotein, Nab2p. Dissecting the requirements for Nab2p-facilitated poly(A) RNA export. *J Biol Chem* 278: 6731-6740.
- Masuda S, Das R, Cheng H, Hurt E, Dorman N, Reed R. 2005. Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* 19: 1512-1517.

- Mathews C, Van Holde K. E. , Ahern K. G. 2002. Bioquímica: Pearson Educación.
- Matsumoto K. 2006. [Formation of mRNPs and their remodeling]. *Seikagaku* 78: 404-408.
- Neer EJ, Schmidt CJ, Nambudripad R, Smith TF. 1994. The ancient regulatory-protein family of WD-repeat proteins. *Nature* 371: 297-300.
- Nelson DL, Michael M. Cox. 2000. *Lehninger Principios de Bioquímica*. Barcelona: Ediciones Omega.
- Newman A. 1997. RNA splicing: out of the loop. *Curr Biol* 7: R418-420.
- Niewold TA, van der Meulen J, Kerstens HH, Smits MA, Hulst MM. 2010. Transcriptomics of enterotoxigenic *Escherichia coli* infection. Individual variation in intestinal gene expression correlates with intestinal function. *Vet Microbiol* 141: 110-114.
- Niraula TNK, S., Inoue, M., Kigawa, T., Yokoyama, S. 2005. Solution structure of the death domain of nuclear matrix protein p84: RCSB Protein Data Bank Database 1WXP.
- Olesen JR, Libri D, Jensen TH. 2005. A link between transcription and mRNP quality in *Saccharomyces cerevisiae*. *RNA Biol* 2: 45-48.
- Oppizzi M. 2004. Genetic, Biochemical and Electron Microscopic Analysis of Components Involved in Transcription-Coupled mRNA Export. Ruperto-Carola University of Heidelberg, Germany, Heidelberg, Germany.
- Orlova EV, Dube P, Harris JR, Beckman E, Zemlin F, Markl J, van Heel M. 1997. Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *J Mol Biol* 271: 417-437.
- Parenteau J, Durand M, Veronneau S, Lacombe AA, Morin G, Guerin V, Cecez B, Gervais-Bird J, Koh CS, Brunelle D, Wellinger RJ, Chabot B, Abou Elela S. 2008. Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol Biol Cell* 19: 1932-1941.
- Parker R, Sheth U. 2007. P bodies and the control of mRNA translation and degradation. *Mol Cell* 25: 635-646.
- Pascual-Montano A, Donate LE, Valle M, Barcena M, Pascual-Marqui RD, Carazo JM. 2001. A novel neural network technique for analysis and classification of EM single-particle images. *J Struct Biol* 133: 233-245.
- Penczek P, Radermacher M, Frank J. 1992. Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy* 40: 33-53.
- Peng WT, Robinson MD, Mnaimneh S, Krogan NJ, Cagney G, Morris Q, Davierwala AP, Grigull J, Yang X, Zhang W, Mitsakakis N, Ryan OW, Datta N, Jojic V, Pal C, Canadien V, Richards D, Beattie B, Wu LF, Altschuler SJ, Roweis S, Frey BJ, Emili A, Greenblatt JF, Hughes TR. 2003. A panoramic view of yeast noncoding RNA processing. *Cell* 113: 919-933.
- Pervushin K, Riek R, Wider G, Wuthrich K. 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U*

S A 94: 12366-12371.

- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612.
- Pickles LM, Roe SM, Hemingway EJ, Stifani S, Pearl LH. 2002. Crystal structure of the C-terminal WD40 repeat domain of the human Groucho/TLE1 transcriptional corepressor. *Structure* 10: 751-761.
- Piruat JI, Aguilera A. 1996. Mutations in the yeast SRB2 general transcription factor suppress hpr1-induced recombination and show defects in DNA repair. *Genetics* 143: 1533-1542.
- Piruat JI, Aguilera A. 1998. A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *EMBO J* 17: 4859-4872.
- Portman DS, O'Connor JP, Dreyfuss G. 1997. YRA1, an essential *Saccharomyces cerevisiae* gene, encodes a novel nuclear protein with RNA annealing activity. *RNA* 3: 527-537.
- Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. 2005. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21: 3435-3438.
- Proudfoot N. 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* 16: 272-278.
- Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 24: 218-229.
- Rafalska-Metcalf IU, Powers SL, Joo LM, LeRoy G, Janicki SM. 2010. Single cell analysis of transcriptional activation dynamics. *PLoS One* 5: e10272.
- Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res* 12: 1231-1245.
- Realini C, Rogers SW, Rechsteiner M. 1994. KEKE motifs. Proposed roles in protein-protein association and presentation of peptides by MHC class I receptors. *FEBS Lett* 348: 109-113.
- Reed R. 2003. Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol* 15: 326-331.
- Reed R, Hurt E. 2002. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* 108: 523-531.
- Reed R, Cheng H. 2005. TREX, SR proteins and export of mRNA. *Curr Opin Cell Biol* 17: 269-273.
- Rehwinkel J, Herold A, Gari K, Kocher T, Rode M, Ciccarelli FL, Wilm M, Izaurralde E. 2004. Genome-wide analysis of mRNAs regulated by the THO complex in *Drosophila mela-*

nogaster. Nat Struct Mol Biol 11: 558-566.

- Richter K, Nessling M, Lichter P. 2008. Macromolecular crowding and its potential impact on nuclear function. Biochim Biophys Acta 1783: 2100-2107.
- Ritchie DB, Schellenberg MJ, MacMillan AM. 2009. Spliceosome structure: piece by piece. Biochim Biophys Acta 1789: 624-633.
- Rodríguez MS, Dargemont C, Stutz F. 2004. Nuclear export of RNA. Biol Cell 96: 639-655.
- Rondon AG, Garcia-Rubio M, Gonzalez-Barrera S, Aguilera A. 2003a. Molecular evidence for a positive role of Spt4 in transcription elongation. EMBO J 22: 612-620.
- Rondon AG, Jimeno S, Garcia-Rubio M, Aguilera A. 2003b. Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. J Biol Chem 278: 39037-39043.
- Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C, Boulay J, Jensen TH, Stutz F, Devaux F, Libri D. 2008. THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association. Cell 135: 308-321.
- Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5: 725-738.
- Saguez C, Olesen JR, Jensen TH. 2005. Formation of export-competent mRNP: escaping nuclear destruction. Curr Opin Cell Biol 17: 287-293.
- Saguez C, Schmid M, Olesen JR, Ghazy MA, Qu X, Poulsen MB, Nasser T, Moore C, Jensen TH. 2008. Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. Mol Cell 31: 91-103.
- Santos-Rosa H, Aguilera A. 1994. Increase in incidence of chromosome instability and non-conservative recombination between repeats in *Saccharomyces cerevisiae* hpr1 delta strains. Mol Gen Genet 245: 224-236.
- Saxton WO, Koch TL. 1982. Interactive image processing with an off-line minicomputer: organization, performance and applications. J Microsc 127: 69-83.
- Scheres SH, Valle M, Carazo JM. 2005a. Fast maximum-likelihood refinement of electron microscopy images. Bioinformatics 21 Suppl 2: ii243-244.
- Scheres SH, Nunez-Ramirez R, Sorzano CO, Carazo JM, Marabini R. 2008. Image processing for electron microscopy single-particle analysis using XMIPP. Nat Protoc 3: 977-990.
- Scheres SH, Nunez-Ramirez R, Gomez-Llorente Y, San Martin C, Eggermont PP, Carazo JM. 2007. Modeling experimental image formation for likelihood-based classification of electron microscopy data. Structure 15: 1167-1177.
- Scheres SH, Valle M, Nunez R, Sorzano CO, Marabini R, Herman GT, Carazo JM. 2005b. Maximum-likelihood multi-reference refinement for electron microscopy images. J Mol Biol 348: 139-149.
- Schneiter R, Guerra CE, Lampl M, Gogg G, Kohlwein SD, Klein HL. 1999. The *Saccharomyces cerevisiae* hyperrecombination mutant hpr1Delta is synthetically lethal with two

conditional alleles of the acetyl coenzyme A carboxylase gene and causes a defect in nuclear export of polyadenylated RNA. *Mol Cell Biol* 19: 3415-3422.

- Selby CP, Sancar A. 1990. Transcription preferentially inhibits nucleotide excision repair of the template DNA strand in vitro. *J Biol Chem* 265: 21330-21336.
- Shaikh TR, Gao H, Baxter WT, Asturias FJ, Boisset N, Leith A, Frank J. 2008. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat Protoc* 3: 1941-1974.
- Sheets MD, Wickens M. 1989. Two phases in the addition of a poly(A) tail. *Genes Dev* 3: 1401-1412.
- Sims RJ, 3rd, Mandal SS, Reinberg D. 2004. Recent highlights of RNA-polymerase-II-mediated transcription. *Curr Opin Cell Biol* 16: 263-271.
- Sleator RD, Walsh P. 2010. An overview of in silico protein function prediction. *Arch Microbiol* 192: 151-155.
- Smith TF, Gaitatzes C, Saxena K, Neer EJ. 1999. The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* 24: 181-185.
- Smolka MB, Albuquerque CP, Chen SH, Zhou H. 2007. Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc Natl Acad Sci U S A* 104: 10364-10369.
- Sorzano CO, Marabini R, Velazquez-Muriel J, Bilbao-Castro JR, Scheres SH, Carazo JM, Pascual-Montano A. 2004. XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol* 148: 194-204.
- Sorzano CO, Bilbao-Castro JR, Shkolnisky Y, Alcorlo M, Melero R, Caffarena-Fernandez G, Li M, Xu G, Marabini R, Carazo JM. 2010. A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J Struct Biol* 171: 197-206.
- Spirin AS, Nemer M. 1965. Messenger RNA in early sea-urchin embryos: cytoplasmic particles. *Science* 150: 214-217.
- Spohr G, Granboulan N, Morel C, Scherrer K. 1970. Messenger RNA in HeLa cells: an investigation of free and polyribosome-bound cytoplasmic messenger ribonucleoprotein particles by kinetic labelling and electron microscopy. *Eur J Biochem* 17: 296-318.
- Springer TA. 2002. Predicted and experimental structures of integrins and beta-propellers. *Curr Opin Struct Biol* 12: 802-813.
- Stachelska A, Wiczorek Z, Ruszczynska K, Stolarski R, Pietrzak M, Lamphear BJ, Rhoads RE, Darzynkiewicz E, Jankowska-Anyszka M. 2002. Interaction of three *Caenorhabditis elegans* isoforms of translation initiation factor eIF4E with mono- and trimethylated mRNA 5' cap analogues. *Acta Biochim Pol* 49: 671-682.
- Stewart M. 2007. Ratcheting mRNA out of the nucleus. *Mol Cell* 25: 327-330.
- Strasser K, Hurt E. 2000a. Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J* 19: 410-420.

- Strasser K, Hurt E. 2001. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* 413: 648-652.
- Strasser K, Bassler J, Hurt E. 2000b. Binding of the Mex67p/Mtr2p heterodimer to FXFG, GLFG, and FG repeat nucleoporins is essential for nuclear mRNA export. *J Cell Biol* 150: 695-706.
- Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R, Hurt E. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* 417: 304-308.
- Stutz F, Izaurralde E. 2003. The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends Cell Biol* 13: 319-327.
- Stutz F, Bachi A, Doerks T, Braun IC, Seraphin B, Wilm M, Bork P, Izaurralde E. 2000. REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* 6: 638-650.
- Takahashi O, Masuda Y, Muroya A, Furuya T. 2010. [In-silico approaches for fragment-based drug design]. *Yakugaku Zasshi* 130: 349-354.
- Tarn WY, Steitz JA. 1997. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci* 22: 132-137.
- Thomas D, Kuras L, Barbey R, Cherest H, Blaiseau PL, Surdin-Kerjan Y. 1995. Met30p, a yeast transcriptional inhibitor that responds to S-adenosylmethionine, is an essential protein with WD40 repeats. *Mol Cell Biol* 15: 6526-6534.
- Thomsen R, Libri D, Boulay J, Rosbash M, Jensen TH. 2003. Localization of nuclear retained mRNAs in *Saccharomyces cerevisiae*. *RNA* 9: 1049-1057.
- Tsai A, Carstens RP. 2006. An optimized protocol for protein purification in cultured mammalian cells using a tandem affinity purification approach. *Nat Protoc* 1: 2820-2827.
- Valpuesta JM, Henderson R, Frey TG. 1990. Electron cryo-microscopic analysis of crystalline cytochrome oxidase. *J Mol Biol* 214: 237-251.
- Valpuesta JM, Martin-Benito J, Gomez-Puertas P, Carrascosa JL, Willison KR. 2002. Structure and function of a protein folding machine: the eukaryotic cytosolic chaperonin CCT. *FEBS Lett* 529: 11-16.
- Voorma HO, Thomas AA, Van Heugten HA. 1994. Initiation of protein synthesis in eukaryotes. *Mol Biol Rep* 19: 139-145.
- Voynov V, Verstrepen KJ, Jansen A, Runner VM, Buratowski S, Fink GR. 2006. Genes with internal repeats require the THO complex for transcription. *Proc Natl Acad Sci U S A* 103: 14423-14428.
- Wall R. 1973. Physiological gradients in development. A possible role for messenger ribonucleoprotein. *Adv Morphog* 10: 41-114.
- Wang L, Sigworth FJ. 2006a. Cryo-EM and single particles. *Physiology (Bethesda)* 21: 13-18.

- Wang X, Chang Y, Li Y, Zhang X, Goodrich DW. 2006b. Thoc1/Hpr1/p84 is essential for early embryonic development in the mouse. *Mol Cell Biol* 26: 4362-4367.
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189-1191.
- Wellinger RE, Prado F, Aguilera A. 2006. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Mol Cell Biol* 26: 3327-3334.
- West RW, Jr., Kruger B, Thomas S, Ma J, Milgrom E. 2000. RLR1 (THO2), required for expressing lacZ fusions in yeast, is conserved from yeast to humans and is a suppressor of SIN4. *Gene* 243: 195-205.
- Williamson JR. 2009. The ribosome at atomic resolution. *Cell* 139: 1041-1043.
- Windgassen M, Krebber H. 2003. Identification of Gbp2 as a novel poly(A)+ RNA-binding protein involved in the cytoplasmic delivery of messenger RNAs in yeast. *EMBO Rep* 4: 278-283.
- Wise JA. 1993. Guides to the heart of the spliceosome. *Science* 262: 1978-1979.
- Wower J, Wower IK, Zwieb C. 2008. Making the jump: new insights into the mechanism of trans-translation. *J Biol* 7: 17.
- Yang J, Li Y, Khoury T, Alrawi S, Goodrich DW, Tan D. 2008. Relationships of hHpr1/p84/Thoc1 expression to clinicopathologic characteristics and prognosis in non-small cell lung cancer. *Ann Clin Lab Sci* 38: 105-112.
- Yu J, Zhu T, Wang Z, Zhang H, Qian Z, Xu H, Gao B, Wang W, Gu L, Meng J, Wang J, Feng X, Li Y, Yao X, Zhu J. 2007. A novel set of DNA methylation markers in urine sediments for sensitive/specific detection of bladder cancer. *Clin Cancer Res* 13: 7296-7304.
- Zenklusen D, Vinciguerra P, Strahm Y, Stutz F. 2001. The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. *Mol Cell Biol* 21: 4219-4232.
- Zenklusen D, Vinciguerra P, Wyss JC, Stutz F. 2002. Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol Cell Biol* 22: 8241-8253.
- Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9: 40.
- Zhong XY, Wang P, Han J, Rosenfeld MG, Fu XD. 2009. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell* 35: 1-10.
- Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* 419: 182-185.
- Zhu Y, Peterson CL, Christman MF. 1995. HPR1 encodes a global positive regulator of transcription in *Saccharomyces cerevisiae*. *Mol Cell Biol* 15: 1698-1708.

